

70 Rec'd PATENT 30 JUL 1998

FORM PTO-1380 (REV 10-95)

U.S. DEPARTMENT OF COMMERCE PATENT AND TRADEMARK OFFICE

ATTORNEY'S DOCKET NUMBER

**TRANSMITTAL LETTER TO THE UNITED STATES  
DESIGNATED/ELECTED OFFICE (DO/EO/US)  
CONCERNING A FILING UNDER 35 U.S.C. 371**

IVD 913

U. S. APPLICATION NO. (if known, see 37 CFR 1.6)

09/125005

INTERNATIONAL APPLICATION NO.  
PCT/FR97/00214INTERNATIONAL FILING DATE  
February 3, 1997PRIORITY DATE CLAIMED  
February 2, 1996

TITLE OF INVENTION

PURIFIED SR-p70 PROTEIN

APPLICANT(S) FOR DO/EO/US

CAPUT, Daniel, FERRARA, Pascual and KAGHAD, Ahmed Mourad

Applicant herewith submits to the United States Designated/Elected Office (DO/EO/US) the following items and other information:

1. ☒ This is a **FIRST** submission of items concerning a filing under 35 U.S.C. 371.
2. ☐ This is a **SECOND** or **SUBSEQUENT** submission of items concerning a filing under 35 U.S.C. 371.
3. ☒ This express request to begin national examination procedures (35 U.S.C. 371(f)) at any time rather than delay examination until the expiration of the applicable time limit set in 35 U.S.C. 371(b) and PCT Articles 22 and 39(1).
4. ☒ A proper Demand for International Preliminary Examination was made by the 19th month from the earliest claimed priority date.
5. ☒ A copy of the International Application as filed (35 U.S.C. 371(c)(2))
  - a. ☐ is transmitted herewith (required only if not transmitted by the International Bureau).
  - b. ☒ has been transmitted by the International Bureau.
  - c. ☐ is not required, as the application was filed in the United States Receiving Office (RO/US).
6. ☒ A translation of the International Application into English (35 U.S.C. 371 (c)(2)).
7. ☐ Amendments to the claims of the International Application under PCT Article 19 (35 U.S.C. 371 (c)(3))
  - a. ☐ are transmitted herewith (required only if not transmitted by the International Bureau).
  - b. ☐ have been transmitted by the International Bureau.
  - c. ☐ have not been made; however, the time limit for making such amendments has NOT expired.
  - d. ☐ have not been made and will not be made.
8. ☐ A translation of the amendments to the claims under PCT Article 19 (35 U.S.C. 371 (c)(3)).
9. ☒ An oath or declaration of the inventor(s) (35 U.S.C. 371(c)(4)).
10. ☐ A translation of the annexes to the International Preliminary Examination Report under PCT Article 36 (35 U.S.C. 371(c)(5)).

**Items 11. to 16. below concern document(s) or information included:**

11. ☐ An Information Disclosure Statement under 37 CFR 1.97 and 1.98.
12. ☒ An assignment document for recording. A separate cover sheet in compliance with 37 CFR 3.28 and 3.31 is included.
13. ☒ A FIRST preliminary amendment.
  - ☐ A SECOND or SUBSEQUENT preliminary amendment.
14. ☐ A substitute specification.
15. ☐ A change of power of attorney and/or address letter.
16. ☐ Other items or information:



79 Rec'd PCT/PTC 30 JUL 1998

## IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Filing under 35 U.S.C. § 371

Corresponding to International Application Serial No.:  
PCT/FR97/00214Applicants: CAPUT, Daniel, FERRARA, Bernard and  
KAGHAD, Mourad

International Filing Date: February 3, 1997

For: **PURIFIED SR-p70 PROTEIN**

Assistant Commissioner for Patents

Box PCT

Attn: EO/US

Washington, D.C. 20231

Dear Sir:

**CERTIFICATE UNDER 37 C.F.R. 1.10**Express Mail Label Number: EM317281100USDate of Deposit: July 29, 1998I hereby certify that this paper is being deposited with the  
United States Postal Service "Express Mail Post Office to  
Addressee" Service on the date indicated above and is  
addressed to: Asst. Commissioner for Patents, Box PCT,  
Attn: EO/US, Washington, DC 20231.Signature **PRELIMINARY AMENDMENT**

Please amend the above-identified application as follows:

**In the Claims**Please amend Claims 1-36 and add Claims 37 and 38 as follows before calculating the filing  
fee for the above-identified application:1.(Amended)  $\Delta$  [Purified] purified polypeptide, comprising an amino acid sequence  
selected from the group consisting of:

- a) [the] sequence SEQ ID No. 2;
- b) [the] sequence SEQ ID No. 4;
- c) [the] sequence SEQ ID No. 6;
- d) [the] sequence SEQ ID No. 8;
- e) [the] sequence SEQ ID No. 10;
- f) [the] sequence SEQ ID No. 13;
- g) [the] sequence SEQ ID No. 15;
- h) [the] sequence SEQ ID No. 17;
- i) [the] sequence SEQ ID No. 19; and

- j) any biologically active sequence derived from SEQ ID No. 2, SEQ ID No. 4, SEQ ID No. 6, SEQ ID No. 8, SEQ ID No. 10, SEQ ID No. 13, SEQ ID No. 15, SEQ ID No. 17 or SEQ ID No. 19.

2. (Amended) A [Polypeptide] polypeptide according to Claim 1, [characterized in that it] [comprises] comprising [the] an amino acid sequence selected from the group consisting of SEQ ID No. 6, SEQ ID No. 13, SEQ ID No. 15, SEQ ID No. 17 and SEQ ID No. 19.

3. (Amended) A [Polypeptide] polypeptide according to Claim 1, [characterized in that it comprises] comprising [the] a sequence lying between:

- residue 110 and residue 310 of SEQ ID No. 2 or 6;
- residue 60 and residue 260 of SEQ ID No. 8.

4. (Amended) A [Polypeptide] polypeptide according to Claim 1, [characterized in that it] which [results] is produced from an alternative splicing of [the] messenger RNA of [the] a corresponding gene.

5. (Amended) A [Polypeptide] polypeptide according to [any one of the preceding claims,] Claim 1 [characterized in that it] that is a recombinant polypeptide produced in the form of a fusion protein.

6. (Amended) An [Isolated] isolated nucleic acid sequence coding for a polypeptide according to [any one of the preceding claims] Claim 1.

7. (Amended) An [Isolated] isolated nucleic acid sequence according to Claim 6, [characterized in that it is] said nucleic acid having a sequence selected from the group consisting of:

- a) [the] sequence SEQ ID No. 1;
- b) [the] sequence SEQ ID No. 3;
- c) [the] sequence SEQ ID No. 5;
- d) [the] sequence SEQ ID No. 7;
- e) [the] sequence SEQ ID No. 9;

- f) [the] sequence SEQ ID No. 11;
- g) [the] sequence SEQ ID No. 12;
- h) [the] sequence SEQ ID No. 14;
- i) [the] sequence SEQ ID No. 16;
- j) [the] sequence SEQ ID No. 18;
- k) [the] nucleic acid sequences capable of hybridizing specifically with [the] sequence SEQ ID No. 1, SEQ ID No. 3, SEQ ID No. 5, SEQ ID No. 7, SEQ ID No. 9, SEQ ID No. 11, SEQ ID No. 12, SEQ ID No. 14, SEQ ID No. 16 or SEQ ID No. 18 or with [the] sequences complementary to them, or of hybridizing specifically with their proximal sequences; and
- l) [the] sequences derived from the sequences a), b), c), d), e), f), g), h), i), j) or k) as a result of the degeneracy of the genetic code, mutation, deletion, insertion, and alternative splicing or an allelic variability.

8. (Amended) Δ [Nucleotide] nucleotide sequence according to Claim 6, [characterized in that it is a sequence] selected from the group consisting of SEQ ID No. 5, SEQ ID No. 12, SEQ ID No. 14, SEQ ID No. 16 and SEQ ID No. 18 and coding, respectively, for the polypeptide of sequences SEQ ID No. 6, SEQ ID No. 13, SEQ ID No. 15, SEQ ID No. 17 and SEQ ID No. 19.

9. (Amended) Δ [Cloning] cloning and/or expression vector containing a nucleic acid sequence according to [any one of Claims] Claim 6 [to 8].

10. (Amended) Δ [Vector] vector, according to Claim 9, [characterized in that it] which is [the] plasmid pSE1.

11. (Amended) Δ [Host] host cell transfected by a vector according to Claim 9 [or 10].

12. (Amended) Δ [Transfected] transfected host cell, according to Claim 11, [characterized in that it] which is *E. coli* MC 1061.

13. (Amended) Δ [Nucleotide] nucleotide probe or nucleotide primer[, characterized in that it] which hybridizes specifically with [any one of the sequences according to Claims] the nucleic acid of Claim 6 [to 8] or [the] a nucleic acid having sequences complementary to them or [the corresponding] messenger RNAs corresponding to them or [the corresponding] genes corresponding to them.

14. (Amended) Δ [Probe] probe or primer according to Claim 13[, characterized in] that [it] contains at least 16 nucleotides.

15. (Amended) Δ [Probe] probe or primer according to Claim 13 [characterized in that it] that comprises the whole of the sequence of the gene coding for [one of the polypeptides of Claim 1] a polypeptide, wherein said polypeptide comprises an amino acid sequence selected from the group consisting of:

- a) sequence SEQ ID No. 2;
- b) sequence SEQ ID No. 4;
- c) sequence SEQ ID No. 6;
- d) sequence SEQ ID No. 8;
- e) sequence SEQ ID No. 10;
- f) sequence SEQ ID No. 13;
- g) sequence SEQ ID No. 15;
- h) sequence SEQ ID No. 17;
- i) sequence SEQ ID No. 19; and
- j) any biologically active sequence derived from SEQ ID No. 2, SEQ ID No. 4, SEQ ID No. 6, SEQ ID No. 8, SEQ ID No. 10, SEQ ID No. 13, SEQ ID No. 15, SEQ ID No. 17 or SEQ ID No. 19.

16.(Amended) Δ [Nucleotide] nucleotide probe or primer selected from the group consisting of the following oligonucleotides or sequences complementary to them:

SEQ ID No. 20: GCG AGC TGC CCT CGG AG  
 SEQ ID No. 21: GGT TCT GCA GGT GAC TCA G  
 SEQ ID No. 22: GCC ATG CCT GTC TAC AAG  
 SEQ ID No. 23: ACC AGC TGG TTG ACG GAG

SEQ ID No. 24: GTC AAC CAG CTG GTG GGC CAG  
 SEQ ID No. 25: GTG GAT CTC GGC CTC C  
 SEQ ID No. 26: AGG CCG GCG TGG GGA AG  
 SEQ ID No. 27: CTT GGC GAT CTG GCA GTA G  
 SEQ ID No. 28: GCG GCC ACG ACC GTG AC  
 SEQ ID No. 29: GGC AGC TTG GGT CTC TGG  
 SEQ ID No. 30: CTG TAC GTC GGT GAC CCC  
 SEQ ID No. 31: TCA GTG GAT CTC GGC CTC  
 SEQ ID No. 32: AGG GGA CGC AGC GAA ACC  
 SEQ ID No. 33: CCA TCA GCT CCA GGC TCT C  
 SEQ ID No. 34: CCA GGA CAG GCG CAG ATG  
 SEQ ID No. 35: GAT GAG GTG GCT GGC TGG A  
 SEQ ID No. 36: TGG TCA GGT TCT GCA GGT G  
 SEQ ID No. 37: CAC CTA CTC CAG GGA TGC  
 SEQ ID No. 38: AGG AAA ATA GAA GCG TCA GTC  
 SEQ ID No. 39: CAG GCC CAC TTG CCT GCC  
 and SEQ ID No. 40: CTG TCC CCA AGC TGA TGA G

17. (Amended) The [Use] use of a sequence according to [any one of Claims] Claim 6 [to 8,] for the manufacture of oligonucleotide primers for sequencing reactions or specific amplification reactions according to the PCR technique or any variant of the latter.

18. (Amended) A [Nucleotide] nucleotide primer pair[, characterized in that it comprises] comprising [the] primers selected from the group consisting of the following sequences:

- a) sense primer: GCG AGC TGC CCT CGG AG (SEQ ID No. 20)  
 antisense primer: GGT TCT GCA GGT GAC TCA G (SEQ ID No. 21)
- b) sense primer: GCC ATG CCT GTC TAC AAG (SEQ ID No. 22)  
 antisense primer: ACC AGC TGG TTG ACG GAG (SEQ ID No. 23)
- c) sense primer: GTC AAC CAG CTG GTG GGC CAG (SEQ ID No. 24)  
 antisense primer: GTG GAT CTC GGC CTC C (SEQ ID No. 25)
- d) sense primer: AGG CCG GCG TGG GGA AG (SEQ ID No. 26)

- e) sense primer: CTT GGC GAT CTG GCA GTA G (SEQ ID No. 27)  
antisense primer: GCG GCC ACG ACC GTG A (SEQ ID No. 28)
- f) sense primer: GGC AGC TTG GGT CTC TGG (SEQ ID No. 29)  
antisense primer: CTG TAC GTC GGT GAC CCC (SEQ ID No. 30)
- g) sense primer: TCA GTG GAT CTC GGC CTC (SEQ ID No. 31)  
antisense primer: AGG GGA CGC AGC GAA ACC (SEQ ID No. 32)
- h) sense primer: GGC AGC TTG GGT CTC TGG (SEQ ID No. 29)  
antisense primer: CCA TCA GCT CCA GGC TCT C (SEQ ID No. 33)
- i) sense primer: CCCCCCCCCCCCCCN (where N equals G, A or T)  
antisense primer: CCA GGA CAG GCG CAG ATG (SEQ ID No. 34)
- j) sense primer: CCCCCCCCCCCCCCN (where N equals G, A or T)  
antisense primer: CTT GGC GAT CTG GCA GTA G (SEQ ID No. 27)
- k) sense primer: CAC CTA CTC CAG GGA TGC (SEQ ID No. 37)  
antisense primer: AGG AAA ATA GAA GCG TCA GTC (SEQ ID No. 38) and
- l) sense primer: CAG GCC CAC TTG CCT GCC (SEQ ID No. 39)  
antisense primer: CTG TCC CCA AGC TGA TGA G (SEQ ID No. 40)

19. (Amended) The [Use] use of a sequence according to [any one of Claims] Claim 6 [to 8,] [which is usable] in gene therapy.

20. (Amended) The [Use] use of a sequence according to [any one of Claims] Claim 6 [to 8,] for the production of diagnostic nucleotide probes or primers, or of antisense sequences which are usable in gene therapy.

21. (Amended) The [Use] use of nucleotide primers according to [any one of Claims] Claim 6 [to 8,] for sequencing.

22. (Amended) The [Use] use of a probe or primer according to [any one of Claims] Claim 13 [to 16,] as an *in vitro* diagnostic tool for the detection, by hybridization experiments, of nucleic acid sequences coding for a polypeptide, wherein said polypeptide comprises an amino acid sequence selected from the group consisting of:

- a) sequence SEQ ID No. 2;
- b) sequence SEQ ID No. 4;
- c) sequence SEQ ID No. 6;
- d) sequence SEQ ID No. 8;
- e) sequence SEQ ID No. 10;
- f) sequence SEQ ID No. 13;
- g) sequence SEQ ID No. 15;
- h) sequence SEQ ID No. 17;
- i) sequence SEQ ID No. 19; and
- j) any biologically active sequence derived from SEQ ID No. 2, SEQ ID No. 4, SEQ ID No. 6, SEQ ID No. 8, SEQ ID No. 10, SEQ ID No. 13, SEQ ID No. 15, SEQ ID No. 17 or SEQ ID No. 19 [according to any one of Claims 1 to 4,] in biological samples, or for the demonstration of aberrant syntheses or of genetic abnormalities.

23.(Amended) A [Method] method of *in vitro* diagnosis for the detection of aberrant syntheses or of genetic abnormalities in the nucleic acid sequences coding for a polypeptide, wherein said polypeptide comprises an amino acid sequence selected from the group consisting of:

- a) sequence SEQ ID No. 2;
- b) sequence SEQ ID No. 4;
- c) sequence SEQ ID No. 6;
- d) sequence SEQ ID No. 8;
- e) sequence SEQ ID No. 10;
- f) sequence SEQ ID No. 13;
- g) sequence SEQ ID No. 15;
- h) sequence SEQ ID No. 17;
- i) sequence SEQ ID No. 19; and
- j) any biologically active sequence derived from SEQ ID No. 2, SEQ ID No. 4, SEQ ID No. 6, SEQ ID No. 8, SEQ ID No. 10, SEQ ID No. 13, SEQ ID No. 15, SEQ ID No. 17 or SEQ ID No. 19

[according to any one of Claims 1 to 4, characterized in that it comprises] comprising the steps of:

- [the] bringing of a nucleotide probe according to [any one of Claims] Claim 13 [to 16] into contact with a biological sample under conditions permitting the formation of a hybridization complex between the [said] probe and the [abovementioned] nucleotide sequence, where appropriate after a prior step of amplification of the [abovementioned] nucleotide sequence;
- the detection of the hybridization complex [possibly] formed; and
- where appropriate, [the] sequencing of the hybridization complex' nucleotide sequence [forming the hybridization complex] with the probe of the invention.

24. (Amended) The [Use] use of a nucleic acid sequence according to [any one of Claims] Claim 6 [to 8,] for the production of a recombinant polypeptide wherein said polypeptide comprises an amino acid sequence selected from the group consisting of:

- a) sequence SEQ ID No. 2;
- b) sequence SEQ ID No. 4;
- c) sequence SEQ ID No. 6;
- d) sequence SEQ ID No. 8;
- e) sequence SEQ ID No. 10;
- f) sequence SEQ ID No. 13;
- g) sequence SEQ ID No. 15;
- h) sequence SEQ ID No. 17;
- i) sequence SEQ ID No. 19; and
- j) any biologically active sequence derived from SEQ ID No. 2, SEQ ID No. 4, SEQ ID No. 6, SEQ ID No. 8, SEQ ID No. 10, SEQ ID No. 13, SEQ ID No. 15, SEQ ID No. 17 or SEQ ID No. 19 [according to any one of Claims 1 to 5].

25. (Amended) A [Method] method of production of a recombinant SR-p70 protein, characterized in that transfected cells according to Claim [10 or] 11 are cultured under conditions permitting the expression of a recombinant polypeptide of sequence SEQ ID No. 2, SEQ ID No. 4, SEQ ID No. 6, SEQ ID No. 8, SEQ ID No. 10, SEQ ID No. 13, SEQ ID No. 15,

SEQ ID No. 17 or SEQ ID No. 19 or any biologically active fragment or derivative, and in that the [said] recombinant polypeptide is recovered.

26. (Amended) Mono- or polyclonal antibodies or their fragments, chimeric antibodies or immunoconjugates, characterized in that they are capable of specifically recognizing a polypeptide according to [any one of Claims] Claim 1 [to 4].

27. (Amended) Use of the antibodies according to [the preceding claim,] Claim 26 for the purification or detection of a polypeptide, wherein said polypeptide comprises an amino acid sequence selected from the group consisting of:

- a) sequence SEQ ID No. 2;
- b) sequence SEQ ID No. 4;
- c) sequence SEQ ID No. 6;
- d) sequence SEQ ID No. 8;
- e) sequence SEQ ID No. 10;
- f) sequence SEQ ID No. 13;
- g) sequence SEQ ID No. 15;
- h) sequence SEQ ID No. 17;
- i) sequence SEQ ID No. 19; and
- j) any biologically active sequence derived from SEQ ID No. 2, SEQ ID No. 4, SEQ ID No. 6, SEQ ID No. 8, SEQ ID No. 10, SEQ ID No. 13, SEQ ID No. 15, SEQ ID No. 17 or SEQ ID No. 19 [according to any one of Claims 1 to 4] in a biological sample.

28. (Amended) A [Method] method of *in vitro* diagnosis of pathologies correlated with an expression or an abnormal accumulation of SR-p70 proteins, in particular the phenomena of carcinogenesis, from a biological sample, [characterized in that] comprising the steps of contacting at least one antibody according to Claim 25 [is brought into contact] with the said biological sample under conditions permitting the [possible] formation of specific immunological complexes between an SR-p70 protein and the said antibody or antibodies, and detecting the presence of [in that the] specific immunological complexes [possibly] formed [are detected].

29. (Amended) Δ [Kit] kit for the *in vitro* diagnosis of an expression or an abnormal accumulation of SR-p70 proteins in a biological sample and/or for measuring the level of expression of these proteins in the said sample, comprising:

- at least one antibody according to Claim 25, optionally bound to a support,
- means of visualization of the formation of specific antigen-antibody complexes between an SR-p70 protein and the said antibody, and/or means of quantification of these complexes.

30. (Amended) Δ [Method] method for the early diagnosis of tumour formation, [characterized in that] wherein autoantibodies directed against an SR-p70 protein are demonstrated in a serum sample drawn from an individual, according to the steps that [consist in] comprise bringing a serum sample drawn from an individual into contact with a polypeptide of the invention, optionally bound to a support, under conditions permitting the formation of specific immunological complexes between the said polypeptide and [the] autoantibodies [possibly] present in the serum sample, and in that the specific immunological complexes [possibly] formed are detected.

31. (Amended) Δ [Method] method of determination of an allelic variability, a mutation, a deletion, an insertion, a loss of heterozygosity or a genetic abnormality of the SR-p70 gene, characterized in that it utilizes at least one nucleotide sequence according to [any one of Claims] Claim 6 [to 8].

32. (Amended) Δ [Method] method of determination of an allelic variability of the SR-p70 gene at position -30 and -20 relative to the initiation ATG of exon 2 which may be involved in pathologies[, and characterized in that it comprises at least] comprising:

- a step during which exon 2 of the SR-p70 gene carrying the target sequence is amplified by PCR using a pair of oligonucleotide primers according to [any one of Claims] Claim 6 [to 8];
- a step during which the amplified products are treated with a restriction enzyme whose cleavage site corresponds to the allele sought and;

- a step during which at least one of the products of the enzyme reaction is detected or assayed.

33. (Amended) Δ [Pharmaceutical] pharmaceutical composition comprising an effective amount of [as active principle a] the polypeptide according to [any one of Claims] Claim 1 [to 4].

34. (Amended) Δ [Pharmaceutical] pharmaceutical composition according to [the preceding claim, characterized in that it comprises] Claim 33, comprising a polypeptide comprising an amino acid sequence selected from SEQ ID No. 6, SEQ ID No. 13, SEQ ID No. 15, SEQ ID No. 17 and SEQ ID No. 19.

35. (Amended) Δ [Pharmaceutical] pharmaceutical composition containing an inhibitor or an activator of SR-p70 activity.

36. (Amended) Δ [Pharmaceutical] pharmaceutical composition containing a polypeptide derived from a polypeptide according to [any one of Claims] Claim 1 [to 5, characterized in that it] which is an inhibitor or an activator of SR-p70.

Please add the following new claims.

37. (New) The use of a probe or primer according to Claim 16 as an *in vitro* diagnostic tool for the detection, by hybridization experiments, of nucleic acid sequences coding for a polypeptide, wherein said polypeptide comprises an amino acid sequence selected from the group consisting of:

- a) sequence SEQ ID No. 2;
- b) sequence SEQ ID No. 4;
- c) sequence SEQ ID No. 6;
- d) sequence SEQ ID No. 8;
- e) sequence SEQ ID No. 10;
- f) sequence SEQ ID No. 13;
- g) sequence SEQ ID No. 15;

- h) sequence SEQ ID No. 17;
- i) sequence SEQ ID No. 19; and
- j) any biologically active sequence derived from SEQ ID No. 2, SEQ ID No. 4, SEQ ID No. 6, SEQ ID No. 8, SEQ ID No. 10, SEQ ID No. 13, SEQ ID No. 15, SEQ ID No. 17 or SEQ ID No. 19 in biological samples, or for the demonstration of aberrant syntheses or of genetic abnormalities.

38. (New) A method of *in vitro* diagnosis for the detection of aberrant syntheses or of genetic abnormalities in the nucleic acid sequences coding for a polypeptide, said polypeptide comprising an amino acid sequence selected from the group consisting of:

- a) sequence SEQ ID No. 2;
- b) sequence SEQ ID No. 4;
- c) sequence SEQ ID No. 6;
- d) sequence SEQ ID No. 8;
- e) sequence SEQ ID No. 10;
- f) sequence SEQ ID No. 13;
- g) sequence SEQ ID No. 15;
- h) sequence SEQ ID No. 17;
- i) sequence SEQ ID No. 19; and
- j) any biologically active sequence derived from SEQ ID No. 2, SEQ ID No. 4, SEQ ID No. 6, SEQ ID No. 8, SEQ ID No. 10, SEQ ID No. 13, SEQ ID No. 15, SEQ ID No. 17 or SEQ ID No. 19

comprising the steps of:

- bringing of a nucleotide probe according to Claim 16 into contact with a biological sample under conditions permitting the formation of a hybridization complex between the probe and the nucleotide sequence, where appropriate, after a prior step of amplification of the nucleotide sequence;
- the detection of the hybridization complex formed; and
- where appropriate, sequencing of the hybridization complex' nucleotide sequence with the probe of the invention.

**REMARKS**

Claims 1-36 have been amended in order to limit the multiple dependencies of these claims and to present them in the appropriate U.S. claim format.

New claims 37 and 38 have been added by the foregoing amendments. Support for these amendments can be found, for example, in original claims 22 and 23, wherein the subject matter now claimed is specifically set forth.

Respectfully submitted,

Date: 29 July 1998

May P. Bauman  
May P. Bauman  
Registration No. 31,926

Address:  
Patent Department  
Sanofi Pharmaceuticals, Inc.  
9 Great Valley Parkway  
P.O. Box 3026  
Malvern, PA 19355  
Telephone No. (610) 889-6338  
Facsimile: (610) 889-8799

[illegible]

filed on February 3, 1997  
in the name of SANOFI

SANOFI

PURIFIED SR-p70 PROTEIN

Abstract

The invention relates to new nucleic acid sequences of the family of tumour-suppressing genes related to the gene for the p53 protein, and to the corresponding protein sequences.

The invention relates to new nucleic acid sequences of the family of tumour-suppressing genes related to the gene for the p53 protein, and to the corresponding protein sequences.

5 The invention also relates to the prophylactic, therapeutic and diagnostic applications of these sequences, in particular in the field of pathologies linked to the phenomena of apoptosis or of cell transformation.

10 Tumour-suppressing genes perform a key role in protection against the phenomena of carcinogenesis, and any modification capable of bringing about the loss of one of these genes, its inactivation or its dysfunction may have oncogenic character, thereby creating favourable  
15 conditions for the development of a malignant tumour.

The authors of the present invention have identified transcription products of a new gene, as well as the corresponding proteins. This gene, SR-p70, is related to the p53 tumour-suppressing gene, the  
20 antitumour activity of which is linked to its transcription factor activity, and more specifically to the controls exerted on the activity of the Bax and Bcl-2 genes which are instrumental in the mechanisms of cell death.

25 Hence the present invention relates to purified SR-p70 proteins, or biologically active fragments of the latter.

The invention also relates to isolated nucleic acid sequences coding for the said proteins or their  
30 biologically active fragments, and to specific oligonucleotides obtained from these sequences.

It relates, in addition, to the cloning and/or expression vectors containing at least one of the nucleotide sequences defined above, and the host cells  
35 transfected by these cloning and/or expression vectors under conditions permitting the replication and/or expression of one of the said nucleotide sequences.

The methods of production of recombinant SR-p70 proteins or their biologically active fragments by the  
40 transfected host cells also form part of the invention.

The invention also comprises antibodies or antibody derivatives specific for the proteins defined above.

5 It relates, in addition, to methods of detection of cancers, either by measuring the accumulation of SR-p70 proteins in the tumours according to immunohistochemical techniques, or by demonstrating autoantibodies directed against these proteins in patients' serum.

10 The invention also relates to any inhibitor or activator of SR-p70 activity, for example of protein-protein interaction, involving SR-p70.

15 It also relates to antisense oligonucleotide sequences specific for the above nucleic acid sequences, capable of modulating *in vivo* the expression of the SR-p70 gene.

20 Lastly, the invention comprises a method of gene therapy, in which vectors such as, for example, inactivated viral vectors capable of transferring coding sequences for a protein according to the invention are injected into cells deficient for this protein, for purposes of regulating the phenomena of apoptosis or of reversion of transformation.

25 A subject of the present invention is a purified polypeptide comprising an amino acid sequence selected from:

- a) the sequence SEQ ID No. 2;
- b) the sequence SEQ ID No. 4;
- c) the sequence SEQ ID No. 6;
- 30 d) the sequence SEQ ID No. 8;
- e) the sequence SEQ ID No. 10;
- f) the sequence SEQ ID No. 13;
- g) the sequence SEQ ID No. 15;
- h) the sequence SEQ ID No. 17;
- 35 i) the sequence SEQ ID No. 19;
- j) any biologically active sequence derived from SEQ ID No. 2, SEQ ID No. 4, SEQ ID No. 6, SEQ ID No. 8, SEQ ID No. 10, SEQ ID No. 13, SEQ ID No. 15, SEQ ID No. 17 or SEQ ID No. 19.

In the description of the invention, the following definitions are used:

- SR-p70 protein: a polypeptide comprising an amino acid sequence selected from the sequences SEQ ID No. 2, SEQ ID No. 4, SEQ ID No. 6, SEQ ID No. 8, SEQ ID No. 10, SEQ ID No. 13, SEQ ID No. 15, SEQ ID No. 17 or SEQ ID No. 19, or any biologically active fragment or derivative of this polypeptide;

- derivative: any variant polypeptide of the polypeptide of sequence SEQ ID No. 2, SEQ ID No. 4, SEQ ID No. 6, SEQ ID No. 8, SEQ ID No. 10, SEQ ID No. 13, SEQ ID No. 15, SEQ ID No. 17 or SEQ ID No. 19, or any molecule resulting from a modification of a genetic and/or chemical nature of the sequence SEQ ID No. 2, SEQ ID No. 4, SEQ ID No. 6, SEQ ID No. 8, SEQ ID No. 10, SEQ ID No. 13, SEQ ID No. 15, SEQ ID No. 17 or SEQ ID No. 19, that is to say obtained by mutation, deletion, addition, substitution and/or chemical modification of a single amino acid or of a limited number of amino acids, as well as any isoform sequence, that is to say sequence identical to the sequence SEQ ID No. 2, SEQ ID No. 4, SEQ ID No. 6, SEQ ID No. 8, SEQ ID No. 10, SEQ ID No. 13, SEQ ID No. 15, SEQ ID No. 17 or SEQ ID No. 19, or to one of its fragments or modified sequences, containing one or more amino acids in the form of the D enantiomer, the said variant, modified or isoform sequences having retained at least one of the properties that make them biologically active;

- biologically active: capable of binding to DNA and/or of exerting transcription factor activity and/or of participating in the control of the cell cycle, of differentiation and of apoptosis and/or capable of being recognized by the antibodies specific for the polypeptide of sequence SEQ ID No. 2, SEQ ID No. 4, SEQ ID No. 6, SEQ ID No. 8, SEQ ID No. 10, SEQ ID No. 13, SEQ ID No. 15, SEQ ID No. 17 or SEQ ID No. 19, and/or capable of inducing antibodies which recognize this polypeptide.

The manufacture of derivatives may have different objectives, including especially that of increasing the

affinity of the polypeptide for DNA or its transcription factor activity, and that of improving its levels of production, of increasing its resistance to proteases, of modifying its biological activities or of endowing it with new pharmaceutical and/or biological properties.

Among the polypeptides of the invention, the polypeptide of human origin comprising the sequence SEQ ID No. 6, SEQ ID No. 13, SEQ ID No. 15, SEQ ID No. 17 or SEQ ID No. 19 is preferred. The polypeptide of 636 amino acids corresponding to the sequence SEQ ID No. 6 is more than 97% identical to the polypeptide of sequence SEQ ID No. 2.

The polypeptide of sequence SEQ ID No. 2 and that of sequence SEQ ID No. 4 are two expression products of the same gene, and the same applies to the sequences SEQ ID No. 8 and SEQ ID No. 10 and to the sequences SEQ ID No. 6, SEQ ID No. 13, SEQ ID No. 15, SEQ ID No. 17 or SEQ ID No. 19.

As will be explained in the examples, the polypeptide of sequence SEQ ID No. 4 corresponds to a premature termination of the peptide of sequence SEQ ID No. 2, linked to an alternative splicing of the longer transcript (messenger RNA), coding for the polypeptide of SEQ ID No. 2, of the corresponding gene. Similarly, in humans, the polypeptides corresponding to the sequences SEQ ID No. 6, SEQ ID No. 13, SEQ ID No. 15, SEQ ID No. 17 and SEQ ID No. 19, diverge in their composition in respect of the N- and/or C-terminal portions, this being the outcome of alternative splicing of the same primary transcript. The N-terminal peptide sequence of the sequence SEQ ID No. 10 is deleted, this being linked to an alternative splicing of its coding transcript.

Advantageously, the invention relates to a polypeptide corresponding to the DNA binding domain of one of the above polypeptides.

This domain corresponds to the sequence lying between residue 110 and residue 310 for the sequences SEQ ID No. 2 or 6, and between residue 60 and residue 260 for the sequence SEQ ID No. 8.

A subject of the present invention is also nucleic acid sequences coding for a SR-p70 protein or biologically active fragments or derivatives of the latter.

5 More preferably, a subject of the invention is an isolated nucleic acid sequence selected from:

- a) the sequence SEQ ID No. 1;
- b) the sequence SEQ ID No. 3;
- c) the sequence SEQ ID No. 5;
- 10 d) the sequence SEQ ID No. 7;
- e) the sequence SEQ ID No. 9;
- f) the sequence SEQ ID No. 11;
- g) the sequence SEQ ID No. 12;
- h) the sequence SEQ ID No. 14;
- 15 i) the sequence SEQ ID No. 16;
- j) the sequence SEQ ID No. 18;
- k) the nucleic acid sequences capable of hybridizing specifically with the sequence SEQ ID No. 1, SEQ ID No. 3, SEQ ID No. 5, SEQ ID No. 7, SEQ ID No. 9, SEQ ID No. 11, SEQ ID No. 12, SEQ ID No. 14, SEQ ID No. 16 or SEQ ID No. 18 or with the sequences complementary to them, or of hybridizing specifically with their proximal sequences;

- l) the sequences derived from the sequences a), b), c), d), e), f), g), h), i), j) or k) as a result of the degeneracy of the genetic code.

According to a preferred embodiment, a subject of the invention is nucleotide sequences SEQ ID No. 5, SEQ ID No. 12, SEQ ID No. 14, SEQ ID No. 16 and SEQ ID No. 18, corresponding, respectively, to the cDNAs of the human proteins of the sequences SEQ ID No. 6, SEQ ID No. 13, SEQ ID No. 15, SEQ ID No. 17 and SEQ ID No. 19.

The different nucleotide sequences of the invention may be of artificial origin or otherwise. They can be DNA or RNA sequences obtained by the screening of libraries of sequences by means of probes prepared on the basis of the sequences SEQ ID No. 1, 3, 5, 7, 9, 11, 12, 14, 16 or 18. Such libraries may be prepared by traditional techniques of molecular biology which are

known to a person skilled in the art.

The nucleotide sequences according to the invention may also be prepared by chemical synthesis, or alternatively by mixed methods including the chemical or enzymatic modification of sequences obtained by the screening of libraries.

These nucleotide sequences enable nucleotide probes to be produced which are capable of hybridizing strongly and specifically with a nucleic acid sequence, of a genomic DNA or of a messenger RNA, coding for a polypeptide according to the invention or a biologically active fragment of the latter. Such probes also form part of the invention. They may be used as an in vitro diagnostic tool for the detection, by hybridization experiments, of transcripts specific for the polypeptides of the invention in biological samples, or for the demonstration of aberrant syntheses or of genetic abnormalities such as loss of heterozygosity or genetic rearrangement resulting from a polymorphism, from mutations or from a different splicing.

The probes of the invention contain at least 10 nucleotides, and contain at most the whole of the sequence of the SR-p70 gene or of its cDNA contained, for example, in a cosmid.

Among the shortest probes, that is to say of approximately 10 to 20 nucleotides, the appropriate hybridization conditions correspond to the stringent conditions normally used by a person skilled in the art.

The temperature used is preferably between  $T_m - 5^\circ\text{C}$  and  $T_m - 30^\circ\text{C}$ , and as a further preference between  $T_m - 5^\circ\text{C}$  and  $T_m - 10^\circ\text{C}$ ,  $T_m$  being the melting temperature, the temperature at which 50% of the paired DNA strands separate.

The hybridization is preferably conducted in solutions of high ionic strength, such as, in particular,  $6 \times \text{SSC}$  solutions.

Advantageously, the hybridization conditions used are as follows:

- temperature:  $42^\circ\text{C}$ ,

- hybridization buffer: 6 x SSC, 5 x Denhart's, 0.1% SDS, as described in Example III.

Advantageously, these probes are represented by the following oligonucleotides or the sequences

5 complementary to them:

SEQ ID No. 20: GCG AGC TGC CCT CGG AG  
SEQ ID No. 21: GGT TCT GCA GGT GAC TCA G  
SEQ ID No. 22: GCC ATG CCT GTC TAC AAG  
SEQ ID No. 23: ACC AGC TGG TTG ACG GAG  
10 SEQ ID No. 24: GTC AAC CAG CTG GTG GGC CAG  
SEQ ID No. 25: GTG GAT CTC GGC CTC C  
SEQ ID No. 26: AGG CCG GCG TGG GGA AG  
SEQ ID No. 27: CTT GGC GAT CTG GCA GTA G  
SEQ ID No. 28: GCG GCC ACG ACC GTG AC  
15 SEQ ID No. 29: GGC AGC TTG GGT CTC TGG  
SEQ ID No. 30: CTG TAC GTC GGT GAC CCC  
SEQ ID No. 31: TCA GTG GAT CTC GGC CTC  
SEQ ID No. 32: AGG GGA CGC AGC GAA ACC  
SEQ ID No. 33: CCA TCA GCT CCA GGC TCT C  
20 SEQ ID No. 34: CCA GGA CAG GCG CAG ATG  
SEQ ID No. 35: GAT GAG GTG GCT GGC TGG A  
SEQ ID No. 36: TGG TCA GGT TCT GCA GGT G  
SEQ ID No. 37: CAC CTA CTC CAG GGA TGC  
SEQ ID No. 38: AGG AAA ATA GAA GCG TCA GTC  
25 SEQ ID No. 39: CAG GCC CAC TTG CCT GCC  
SEQ ID No. 40: CTG TCC CCA AGC TGA TGA G

Preferably, the probes of the invention are labelled prior to their use. To this end, several techniques are within the capacity of a person skilled in  
30 the art (fluorescent, radioactive, chemoluminescence, enzyme, and the like, labelling).

The in vitro diagnostic methods in which these nucleotide probes are employed are included in the subject of the present invention.

35 These methods relate, for example, to the detection of abnormal syntheses (e.g. accumulation of transcription products) or of genetic abnormalities, such as loss of heterozygosity and genetic rearrangement, and point mutations in the nucleotide sequences of nucleic

acids coding for an SR-p70 protein, according to the definition given above.

The nucleotide sequences of the invention are also useful for the manufacture and use of  
5 oligonucleotide primers for sequencing reactions or specific amplification reactions according to the so-called PCR technique or any variant of the latter (ligase chain reaction (LCR), etc).

Preferred primer pairs consist of primers  
10 selected from the nucleotide sequences: SEQ ID No. 1: monkey sequence of 2,874 nucleotides, and SEQ ID No. 5: human SR-p70a cDNA, in particular upstream of the ATG translation initiation codon and downstream of the TGA translation stop codon.

15 Advantageously, these primers are represented by the following pairs:

- pair No. 1:

sense primer: GCG AGC TGC CCT CGG AG (SEQ ID No. 20)

antisense primer: GGT TCT GCA GGT GAC TCA G (SEQ ID No. 21)

20

- pair No. 2:

sense primer: GCC ATG CCT GTC TAC AAG (SEQ ID No. 22)

antisense primer: ACC AGC TGG TTG ACG GAG (SEQ ID No. 23)

- pair No. 3:

25

sense primer: GTC AAC CAG CTG GTG GGC CAG (SEQ ID No. 24)

antisense primer: GTG GAT CTC GGC CTC C (SEQ ID No. 25)

- pair No. 4:

30

sense primer: AGG CCG GCG TGG GGA AG (SEQ ID No. 26)

antisense primer: CTT GGC GAT CTG GCA GTA G (SEQ ID No. 27)

- pair No. 5:

sense primer: GCG GCC ACG ACC GTG A (SEQ ID No. 28)

antisense primer: GGC AGC TTG GGT CTC TGG (SEQ ID No. 29)

- pair No. 6:

35

sense primer: CTG TAC GTC GGT GAC CCC (SEQ ID No. 30)

antisense primer: TCA GTG GAT CTC GGC CTC (SEQ ID No. 31)

- pair No. 7:

sense primer: AGG GGA CGC AGC GAA ACC (SEQ ID No. 32)

antisense primer: GGC AGC TTG GGT CTC TGG (SEQ ID No. 29)

- pair No. 8:

sense primer: CCCCCCCCCCCCCCN (where N equals G, A or T)  
antisense primer: CCA TCA GCT CCA GGC TCT C (SEQ ID No. 33)

- pair No. 9:

5 sense primer: CCCCCCCCCCCCCCN (where N equals G, A or T)  
antisense primer: CCA GGA CAG GCG CAG ATG (SEQ ID No. 34)

- pair No. 10:

sense primer: CCCCCCCCCCCCCCN (where N equals G, A or T)  
antisense primer: CTT GGC GAT CTG GCA GTA G (SEQ ID No. 27)

10 - pair No. 11:

sense primer: CAC CTA CTC CAG GGA TGC (SEQ ID No. 37)  
antisense primer: AGG AAA ATA GAA GCG TCA GTC (SEQ ID No. 38)

- pair No. 12:

15 sense primer: CAG GCC CAC TTG CCT GCC (SEQ ID No. 39)  
antisense primer: CTG TCC CCA AGC TGA TGA G (SEQ ID No. 40)

These primers correspond to the sequences  
extending, respectively:

- 20 - from nucleotide No. 124 to nucleotide No. 140  
on SEQ ID No. 1 and from nucleotide No. 1 to  
nucleotide No. 17 on SEQ ID No. 5 for SEQ ID  
No. 20
- 25 - from nucleotide No. 2280 to nucleotide No. 2262  
on SEQ ID No. 1 and from nucleotide No. 2156 to  
nucleotide 2138 on SEQ ID No. 5 for SEQ ID No.  
21
- from nucleotide No. 684 to nucleotide No. 701  
on SEQ ID No. 1 for SEQ ID No. 22
- from nucleotide No. 1447 to nucleotide No. 1430  
on SEQ ID No. 1 and from nucleotide 1324 to  
30 nucleotide 1307 on SEQ ID No. 5 for SEQ ID No.  
23
- from nucleotide 1434 to nucleotide 1454 on SEQ  
ID No. 1 and from nucleotide 1311 to nucleotide  
1331 on SEQ ID No. 5 for SEQ ID No. 24
- 35 - from nucleotide 2066 to nucleotide 2051 on SEQ  
ID No. 1 and from nucleotide 1940 to nucleotide

- 1925 on SEQ ID No. 5 for SEQ ID No. 25
- from nucleotide 16 to nucleotide 32 on SEQ ID No. 5 for SEQ ID No. 26
  - from nucleotide 503 to nucleotide 485 on SEQ ID No. 5 for SEQ ID No. 27
  - from nucleotide 160 to nucleotide 176 on SEQ ID No. 11 for SEQ ID No. 28
  - from nucleotide 1993 to nucleotide 1976 on SEQ ID No. 5 for SEQ ID No. 29
  - from nucleotide 263 to nucleotide 280 on SEQ ID No. 11 for SEQ ID No. 30
  - from nucleotide 1943 to nucleotide 1926 on SEQ ID No. 5 for SEQ ID No. 31
  - from nucleotide 128 to nucleotide 145 on the nucleotide sequence depicted in Figure 22 for SEQ ID No. 32
  - from nucleotide 1167 to nucleotide 1149 on SEQ ID No. 5 for SEQ ID No. 33
  - from nucleotide 928 to nucleotide 911 on SEQ ID No. 5 for SEQ ID No. 34
  - from nucleotide 677 to nucleotide 659 on SEQ ID No. 5 for SEQ ID No. 35
  - from nucleotide 1605 to nucleotide 1587 on SEQ ID No. 5 for SEQ ID No. 36
  - from nucleotide 1 to nucleotide 18 on the nucleotide sequence depicted in Figure 13 for SEQ ID No. 37
  - from nucleotide 833 to nucleotide 813 on the nucleotide sequence depicted in Figure 13 for SEQ ID No. 38
  - from nucleotide 25 to nucleotide 42 on the nucleotide sequence depicted in Figure 13 for SEQ ID No. 39
  - from nucleotide 506 to nucleotide 488 on the nucleotide sequence depicted in Figure 13 for SEQ ID No. 40

The nucleotide sequences according to the invention can have, moreover, uses in gene therapy, in particular for controlling the phenomena of apoptosis and

of reversion of transformation.

The nucleotide sequences according to the invention may, moreover, be used for the production of recombinant SR-p70 proteins, according to the definition  
5 which has been given to this term.

These proteins may be produced from the nucleotide sequences defined above, according to techniques of production of recombinant products which are known to a person skilled in the art. In this case,  
10 the nucleotide sequence used is placed under the control of signals permitting its expression in a cell host.

An effective system for production of a recombinant protein necessitates having at one's disposal a vector, for example of plasmid or viral origin, and a  
15 compatible host cell.

The cell host may be selected from prokaryotic systems such as bacteria, or eukaryotic systems such as, for example, yeasts, insect cells, CHO cells (Chinese hamster ovary cells) or any other system advantageously  
20 available. A preferred cell host for the expression of proteins of the invention consists of the *E. coli* bacterium, in particular the strain MC 1061 (Clontec).

The vector must contain a promoter, translation initiation and termination signals and also the  
25 appropriate transcription regulation regions. It must be capable of being maintained stably in the cell and can, where appropriate, possess particular signals specifying the secretion of the translated protein.

These various control signals are selected in accordance with the cell host used. To this end, the nucleotide sequences according to the invention may be inserted into vectors which are autonomously replicating within the selected host, or vectors which are integrative for the chosen host. Such vectors will be  
30 prepared according to methods commonly used by a person skilled in the art, and the clones resulting therefrom may be introduced into a suitable host by standard methods such as, for example, electroporation.

The cloning and/or expression vectors containing

at least one of the nucleotide sequences defined above also form part of the present invention.

A preferred cloning and expression vector is the plasmid pSE1, which contains the elements necessary for its use both as a cloning vector in *E. coli* (origin of replication in *E. coli* and ampicillin resistance gene originating from the plasmid pTZ 18R) and as an expression vector in animal cells (promoter, intron, polyadenylation site, origin of replication of the SV40 virus), as well as the elements enabling it to be copied as a single strand with the object of sequencing (origin of replication of phage fl).

The characteristics of this plasmid are described in Application EP 0,506,574.

Its construction and also the integration of the cDNAs originating from the nucleic acid sequences of the invention are, moreover, described in the examples below.

According to a preferred embodiment, the proteins of the invention are in the form of fusion proteins, in particular in the form of a protein fused with glutathione S-transferase (GST). A designated expression vector in this case is represented by the plasmid vector pGEX-4T-3 (Pharmacia ref-27.4583).

The invention relates, in addition, to the host cells transfected by these aforementioned vectors. These cells may be obtained by introducing into host cells a nucleotide sequence inserted into a vector as defined above, followed by culturing of the said cells under conditions permitting the replication and/or expression of the transfected nucleotide sequence.

These cells are usable in a method of production of a recombinant polypeptide of sequence SEQ ID No. 2, SEQ ID No. 4, SEQ ID No. 6, SEQ ID No. 8, SEQ ID No. 10, SEQ ID No. 12, SEQ ID No. 14, SEQ ID No. 16 or SEQ ID No. 18 or any biologically active fragment or derivative of the latter.

The method of production of a polypeptide of the invention in recombinant form is itself included in the present invention, and is characterized in that the

transfected cells are cultured under conditions permitting the expression of a recombinant polypeptide of sequence SEQ ID No. 2, SEQ ID No. 4, SEQ ID No. 6, SEQ ID No. 8, SEQ ID No. 10, SEQ ID No. 12, SEQ ID No. 14, SEQ ID No. 16 or SEQ ID No. 18 or of any biologically active fragment or derivative of the latter, and in that the said recombinant polypeptide is recovered.

The purification methods used are known to a person skilled in the art. The recombinant polypeptide may be purified from lysates and cell extracts or from the culture medium supernatant, by methods used individually or in combination, such as fractionation, chromatographic methods, immunoaffinity techniques using specific mono- or polyclonal antibodies, and the like. A preferred variant consists in producing a recombinant polypeptide fused to a "carrier" protein (chimeric protein). The advantage of this system is that it permits a stabilization and a decrease in proteolysis of the recombinant product, an increase in solubility during in vitro renaturation and/or a simplification of the purification when the fusion partner possesses an affinity for a specific ligand.

Advantageously, the polypeptides of the invention are fused with glutathione S-transferase at the N-terminal position (Pharmacia "GST" system). The fusion product is, in this case, detected and quantified by means of the enzyme activity of the GST. The colorimetric reagent used is a glutathione acceptor, a substrate for GST. The recombinant product is purified on a chromatographic support to which glutathione molecules have been coupled beforehand.

The mono- or polyclonal antibodies capable of specifically recognizing an SR-p70 protein according to the definition given above also form part of the invention. Polyclonal antibodies may be obtained from the serum of an animal immunized against protein, produced, for example, by genetic recombination according to the method described above, according to standard procedures.

The monoclonal antibodies may be obtained

according to the traditional hybridoma culture method described by Köhler and Milstein, Nature, 1975, 256, 495-497.

Advantageous antibodies are antibodies directed  
5 against the central region lying between residue 110 and residue 310 for the sequences SEQ ID No. 2 or 6, or between residue 60 and residue 260 for the sequence SEQ ID No. 8.

The antibodies according to the invention are,  
10 for example, chimeric antibodies, humanized antibodies or Fab and F(ab')<sub>2</sub> fragments. They may also take the form of immunoconjugates or labelled antibodies.

Moreover, besides their use for the purification  
of the recombinant polypeptides, the antibodies of the  
15 invention, especially the monoclonal antibodies, may also be used for detecting these polypeptides in a biological sample.

Thus they constitute a means of  
immunocytochemical or immunohistochemical analysis of the  
20 expression of SR-p70 proteins on sections of specific tissues, for example by immunofluorescence, gold labelling or enzyme immunoconjugates.

They make it possible, in particular, to  
demonstrate an abnormal accumulation of SR-p70 proteins  
25 in certain tissues or biological samples, which makes them useful for detecting cancers or monitoring the progression or remission of pre-existing cancers.

More generally, the antibodies of the invention  
may be advantageously employed in any situation where the  
30 expression of an SR-p70 protein has to be observed.

Hence the invention also relates to a method of  
in vitro diagnosis of pathologies correlated with an  
expression or an abnormal accumulation of SR-p70 pro-  
teins, in particular the phenomena of carcinogenesis,  
35 from a biological sample, characterized in that at least one antibody of the invention is brought into contact with the said biological sample under conditions permitting the possible formation of specific immuno-logical complexes between an SR-p70 protein and the said

antibody or antibodies, and in that the specific immunological complexes possibly formed are detected.

The invention also relates to a kit for the *in vitro* diagnosis of an abnormal expression or  
5 accumulation of SR-p70 proteins in a biological sample and/or for measuring the level of expression of this protein in the said sample, comprising:

- at least one antibody specific for an SR-p70 protein, optionally bound to a support,
- 10 - means of visualization of the formation of specific antigen-antibody complexes between an SR-p70 protein and the said antibody, and/or means of quantification of these complexes.

The invention also relates to a method of early  
15 diagnosis of tumour formation, by detecting autoantibodies directed against an SR-p70 protein in an individual's serum.

Such a method of early diagnosis is characterized in that a serum sample drawn from an individual is  
20 brought into contact with a polypeptide of the invention, optionally bound to a support, under conditions permitting the formation of specific immunological complexes between the said polypeptide and the autoantibodies possibly present in the serum sample, and  
25 in that the specific immunological complexes possibly formed are detected.

A subject of the invention is also a method of determination of an allelic variability, a mutation, a deletion, an insertion, a loss of heterozygosity or a  
30 genetic abnormality of the SR-p70 gene which may be involved in pathologies, characterized in that it utilizes at least one nucleotide sequence described above. Among the methods of determination of an allelic variability, a mutation, a deletion, an insertion, a loss  
35 of heterozygosity or a genetic abnormality of the SR-p70 gene, preference is given to the method which is characterized in that it comprises at least one step of PCR amplification of the target nucleic acid sequence of SR-p70 liable to exhibit a polymorphism, a mutation, a

deletion or an insertion, using a pair of primers of nucleotide sequences defined above, a step during which the amplified products are treated using a suitable restriction enzyme and a step during which at least one  
5 of the products of the enzyme reaction is detected or assayed.

The invention also comprises pharmaceutical compositions comprising as active principle a polypeptide corresponding to the above definitions, preferably in  
10 soluble form, in combination with a pharmaceutically acceptable vehicle.

Such compositions afford a novel approach to treating the phenomena of carcinogenesis at the level of the control of multiplication and cell differentiation.

15 Preferably, these compositions can be administered systemically, preferably intravenously, intramuscularly, intradermally or orally.

Their optimal modes of administration, dosages and pharmaceutical dosage forms may be determined  
20 according to the criteria generally borne in mind in establishing a therapeutic treatment suitable for a patient, such as, for example, the patient's age or body weight, the severity of his or her general state, the tolerability of treatment and the observed side effects,  
25 and the like.

Lastly, the invention comprises a method of gene therapy, in which nucleotide sequences coding for an SR-p70 protein are transferred to target cells by means of inactivated viral vectors.

30 Other features and advantages of the invention are to be found in the remainder of the description, with the examples and the figures for which the legends are given below.

#### LEGEND TO THE FIGURES

35 Figure 1: Nucleic acid comparison of monkey SR-p70a cDNA (corresponding to SEQ ID No. 1) with the nucleic acid sequence of monkey p53 cDNA.

- Figure 2: Protein comparison of monkey SR-p70a with monkey p53 protein (sw: p53-cerae).
- Figure 3: Comparison of the nucleic acid sequence of monkey SR-p70a and b cDNA (corresponding, respectively, to SEQ ID No. 1 and SEQ ID No. 3).
- Figure 4: Nucleic acid sequence and deduced protein sequence of monkey SR-p70a.
- Figure 5: Partial nucleic acid sequence and complete deduced protein sequence of monkey SR-p70b.
- Figure 6: Partial nucleic acid sequence and deduced complete protein sequence of human SR-p70a (corresponding to SEQ ID No. 5).
- Figure 7: Partial nucleic acid sequence and complete deduced protein sequence of mouse SR-p70c (corresponding to SEQ ID No. 7).
- Figure 8: Partial nucleic acid sequence and partially deduced protein sequence of mouse SR-p70a (corresponding to SEQ ID No. 9).
- Figure 9: Multialignment of the proteins deduced from monkey (a and b), human (a) and mouse (a and c) SR-p70 cDNAs.
- Figure 10a: Immunoblot of the SR-p70 protein.
- Figure 10b: Detection of the endogenous SR-p70 protein.
- Figure 11: Chromosomal localization of the human SR-p70 gene. The signal appears on chromosome 1, in the p36 region.
- Figure 12: Genomic structure of the SR-p70 gene and

5 comparison with that of the p53 gene. The human protein sequences of SR-p70a (upper line of the alignment) and of p53 (lower line) are divided up into peptides on the basis of the respective exons from which they are encoded. The figures beside the arrows correspond to the numbering of the corresponding exons.

10 Figure 13: Human genomic sequence of SR-p70 from the 3' end of intron 1 to the 5' end of exon 3. The introns are boxed. At positions 123 and 133, two variable nucleic acid positions are localized (G → A at 123 and C → T at 133). The restriction sites for the enzyme StyI are underlined (position 130 in the case where a T is present instead of a C at position 133, position 542 and position 610). The arrows indicate the positions of the nucleic acid primers used in Example XI.

20 Figure 14: Nucleic acid comparison of the 5' region of the human cDNAs of SR-p70d and of SR-p70a.

Figure 15: Multialignment of the nucleic acid sequences corresponding to human SR-p70a, b, d, e, and f.

25 Figure 16: Multialignment of the proteins deduced from human SR-p70 (a, b, d, e and f) cDNAs.

30 Figure 17: Partial nucleic acid sequence and partial deduced protein sequence of human SR-p70a. The two bases in bold characters correspond to two variable positions (see Figure 6). This sequence possesses a more complete non-coding 5' region than the one presented in Figure 6.

Figure 18: Analysis of the SR-p70a transcripts after PCR amplification.

lane M: 1 kb ladder (GIBCO-BRL) molecular weight markers

lane 1: line HT29

lane 3: line SK-N-AS

lane 5: line UMR-32

lane 7: line U-373 MG

lane 9: line SW 480

lane 11: line CHP 212

lane 13: line SK-N-MC

lanes 2, 4, 6, 8, 10, 12, 14: negative controls corresponding to lanes 1, 3, 5, 7, 9, 11 and 13, respectively (absence of inverse transcriptase in the RT-PCR reaction).

Figure 19: A: Analysis by agarose gel electrophoresis of genomic fragments amplified by PCR (from the 3' end of intron 1 to the 5' end of exon 3). The numbering of the lanes corresponds to the numbering of the control population. Lane M: molecular weight markers (1 kb ladder).

B: Analysis identical to that of part A, after digestion of the same samples with the restriction enzyme StyI.

Figure 20: Diagrammatic representation with a partial restriction map of the plasmid pCDNA3 containing human SR-p70a.

## EXAMPLE I

### Cloning of SR-p70 cDNA from COS-3 cells

#### 1. Culturing of COS-3 cells

5 COS-3 cells (African green monkey kidney cells transformed with the SV 40 virus T antigen) are cultured in DMEM medium (GIBCO-BRL reference 41 965-047) containing 2 mM L-glutamine and supplemented with 50 mg/l of gentamicin and 5% of foetal bovine serum (GIBCO-BRL reference 10231-074) to semi-confluence.

#### 10 2. Preparation of the messenger RNA

##### a) Extraction of the messenger RNA

The cells are recovered in the following manner:

15 - the adherent cells are washed twice with PBS buffer (phosphate buffered saline, reference 04104040-GIBCO-BRL), then scraped off with a rubber scraper and centrifuged.

The cell pellet is suspended in the lysis buffer of the following composition: 4 M guanidine thiocyanate; 25 mM sodium citrate pH 7; 0.5% sarcosyl; 0.1 M  $\beta$ -mercaptoethanol. The suspension is sonicated using an Ultra-Turrax No. 231256 sonicator (Janke and Kundel) at maximum power for one minute. Sodium acetate pH 4 is added to a concentration of 0.2 M. The solution is extracted with one volume of 20 a phenol/chloroform (5/1 v/v) mixture. The RNA contained in the aqueous phase is precipitated at -20°C using one volume of isopropanol. The pellet is resuspended in the lysis buffer. The solution is extracted again with a phenol/chloroform mixture and the RNA is precipitated with isopropanol. After washing of the pellet with 70% and then 100% ethanol, the RNA is resuspended in water.

25 b) Purification of the poly(A)<sup>+</sup> fraction of the RNA  
Purification of poly(A)<sup>+</sup> fraction of the RNA is carried out using the DYNAL Dynabeads oligo(dT)<sub>25</sub> kit (reference 610.05) according to the protocol 35

recommended by the manufacturer. The principle is based on the use of superparamagnetic polystyrene beads to which an oligonucleotide poly(dT)<sub>25</sub> is attached. The poly(A)<sup>+</sup> fraction of the RNA is hybridized with the oligo(dT)<sub>25</sub> coupled to the beads, which are trapped on a magnetic support.

3. *Production of the complementary DNA library*

a) *Preparation of the complementary DNA*

From 0.5 µg of the poly(A)<sup>+</sup> RNA from COS-3 cells obtained at the end of step 2, the [<sup>32</sup>P]dCTP-labelled single-stranded complementary DNA is prepared (the complementary DNA obtained possesses a specific activity of 3000 dpm/ng) with the synthetic primer of the following sequence (comprising a BamHI site):

5'<GATCCGGGCC CTTTTTTTT TTT<3'

in a volume of 30 µl of buffer of composition: 50 mM Tris-HCl pH 8.3, 6 mM MgCl<sub>2</sub>, 10 mM DDT, 40 mM KCl, containing 0.5 mM each of the deoxynucleotide triphosphates, 30 µCi of [α-<sup>32</sup>P]dCTP and 30 U of RNasin (Promega). After one hour of incubation at 37°C, then 10 minutes at 50°C, then 10 minutes again at 37°C, with 200 units of the enzyme reverse transcriptase RNase H<sup>-</sup> (GIBCO-BRL reference 8064A), 4 µl of EDTA are added.

b) *Alkaline hydrolysis of the RNA template*

6 µl of 2N NaOH solution are added and the mixture is then incubated for 5 minutes at 65°C.

c) *Purification on a Sephacryl S-400 column*

In order to remove the synthetic primer, the complementary DNA is purified on a column of 1 ml of Sephacryl S-400 (Pharmacia) equilibrated in TE buffer.

The first two radioactive fractions are pooled and precipitated with 1/10 volume of 10 M ammonium acetate solution and 2.5 volumes of ethanol, this being done after extraction with one volume of chloroform.

d) Homopolymer addition of dG

The complementary DNA is elongated at the 3' end with a dG tail with 20 units of the enzyme terminal transferase (Pharmacia 27073001). The mixture is incubated in 20  $\mu$ l of buffer of composition: 30 mM Tris-HCl pH 7.6, 1 mM cobalt chloride, 140 mM cacodylic acid, 0.1 mM DTT, 1 mM dGTP, for 15 minutes at 37°C, and 2  $\mu$ l of 0.5 M EDTA are then added.

10 e) Steps b) and c) are repeated again

f) Pairing of the cloning vector pSE1 (EP 506,574) and the complementary DNA in the presence of the adaptor.

The mixture is centrifuged, the pellet is dissolved in 33  $\mu$ l of TE buffer, 5  $\mu$ l (125 ng) of cloning vector pSE1, 1  $\mu$ l (120 ng) of the adaptor of the following sequence (comprising an ApaI site):

5'AAAAAAAAAAAAAGGGCCCG3'

and 10  $\mu$ l of 200 mM NaCl solution are added, and the reaction mixture is incubated for 5 minutes at 65°C and then allowed to cool to room temperature.

g) Ligation

The cloning vector and the single-stranded cDNA are ligated in a volume of 100  $\mu$ l with 32.5 units of the enzyme phage T4 DNA ligase (Pharmacia reference 270 87002) overnight at 15°C in a buffer of composition: 50 mM Tris-HCl pH 7.5, 10 mM MgCl<sub>2</sub>, 1 mM ATP.

h) Synthesis of the second strand of the cDNA

The proteins are removed by phenol extraction followed by chloroform extraction, and 1/10 volume of 10 mM ammonium acetate solution and then 2.5 volumes of ethanol are then added. The mixture is centrifuged, the pellet is dissolved in a buffer of composition 33 mM Tris-acetate pH 7.9, 62.5 mM potassium acetate, 1 mM magnesium acetate and 1 mM dithiothreitol (DTT), and the second strand of complementary DNA is synthesized in a volume of 30  $\mu$ l with 30 units of the enzyme phage T4 DNA polymerase (Pharmacia reference 270718) and a

mixture of 1 mM the four deoxynucleotide triphosphates dATP, dCTP, dGTP and dTTP as well as two units of phage T4 gene 32 protein (Pharmacia reference 27-0213) for one hour at 37°C. The mixture is extracted with phenol and the traces of phenol are removed with a column of polyacrylamide P10 (Biogel P10-200-400 mesh - reference 15011050 - Biorad).

i) Transformation by electroporation

*E. coli* MC 1061 cells are transformed with the recombinant DNA obtained above by electroporation using a Biorad Gene Pulser apparatus (Biorad) used at 2.5 kV under the conditions specified by the manufacturer, and the bacteria are then grown for one hour in the medium known as LB medium (Sambrook op. cit.) of composition: bactotryptone 10 g/l; yeast extract 5 g/l; NaCl 10 g/l.

The number of independent clones is determined by plating out a 1/1000 dilution of the transformation after the first hour of incubation on a dish of LB medium with the addition of 1.5% of agar (w/v) and 100 µg/ml of ampicillin, hereinafter referred to as LB agar medium. The number of independent clones is 1 million.

j) Analysis of the cDNAs of the library

In the context of the analysis of individual clones of the library by nucleic acid sequencing of the 5' region of the cDNAs, one clone, designated SR-p70a, was shown to exhibit a partial homology with the cDNA of the already known protein, the p53 protein (Genbank X 02469 and X 16384) (Figure 1). The sequences were produced with the United States Biochemical kit (reference 70770) and/or the Applied Biosystems kit (references 401434 and/or 401628), which use the method of Sanger et al., Proc. Natl. Acad. Sci. USA; 1977, 14, 5463-5467. The plasmid DNA is prepared from the WIZARD miniprep kit (Promega reference A7510). The primers used are 16- to 22-mer oligonucleotides, complementary either to

the vector pSE1 in the region immediately at the 5' end of the cDNA, or to the sequence of the cDNA.

A second cDNA was isolated from the same library by screening, in a manner similar to the technique described in EXAMPLE III.3) below, with a fragment of SR-p70a the DNA labelled with <sup>32</sup>P with the BRL "Random Primers DNA labelling systems" kit (reference 18187-013). The hybridization and washing buffers are treated by adding 50% of formamide. The last wash is carried out in 0.1 x SSC/0.1% SDS at 60°C. This second sequence (SR-p70b cDNA) is identical to the first but an internal fragment has been deleted from it (Figure 3).

The two SR-p70 cDNAs, of length 2874 nucleotides (SR-p70a) and 2780 nucleotides (SR-p70b), correspond to the products of a single gene, an alternative splicing bringing about a deletion of 94 bases between nucleotides 1637 and 1732 and a premature termination of the corresponding encoded protein. The proteins deduced from the two cDNAs possess 637 amino acids and 499 amino acids, respectively (Figures 4 and 5).

## EXAMPLE II

Obtaining of the sequence and cloning of the cDNA of the SR-p70a protein from HT-29 (human colon adenocarcinoma) cells

### 1) Culturing of HT-29 cells

The cells are cultured in McCoy's 5 medium (GIBCO 26600-023) with the addition of 10% of foetal calf serum (GIBCO 10081-23) and 50 mg/l of gentamicin, to semi-confluence.

### 2) Preparation of the complementary DNA

The messenger RNA is prepared as described in EXAMPLE I.2. The cDNA is prepared in a manner similar to that described in EXAMPLE I.3, with 5 µg of total messenger RNA, using a poly(T)<sub>12</sub> primer. The reaction is

not interrupted with EDTA.

3) *Specific amplification of the human cDNA by the so-called PCR technique*

The polymerization is carried out with 4  $\mu$ l of  
5 cDNA in 50  $\mu$ l final with the buffer of the following  
composition: 10 mM Tris-HCl pH 8.3, 2.5 mM MgCl<sub>2</sub>, 50 mM  
KCl in the presence of 10% DMSO, 0.5 mM dNTP, 4  $\mu$ g/ml of  
each of the two nucleic acid primers and 2.5 units of TAQ  
DNA polymerase (Boehringer). The primer pairs were  
10 selected on the basis of the nucleic acid sequence of the  
COS-3 SR-p70 clone, in particular upstream of the  
translation initiation ATG and downstream of the  
translation stop TGA, and are of the following  
compositions:

15 sense primer: ACT GGT ACC GCG AGC TGC CCT CGG AG  
Kpn I restriction site

antisense primer: GAC TCT AGA GGT TCT GCA GGT GAC TCA G  
Xba I restriction site

The reaction is carried out for 30 cycles of  
20 94°C/1 minute, 54-60°C/1 minute 30 seconds and 72°C/  
1 minute 30 seconds, followed by a final cycle of  
72°C/6 minutes.

4) *Obtaining of the sequence of the human cDNA*

In a first step, the PCR product is removed from  
25 the oligonucleotides on a column of Sephacryl S-400, and  
then desalted by exclusion chromatography on a column of  
polyacrylamide P10 (Biorad reference 1504144). The  
sequencing reactions are carried out using the Applied  
Biosystems kit (reference 401628) with oligonucleotides  
30 specific for the cDNA. The sequence obtained is very  
similar to that of monkey SR-p70a, and the deduced  
protein contains 636 amino acids (Figure 6).

In a similar manner, other sequences originating  
from human lines or tissues were obtained for the coding

portion of human SR-p70, in particular from the lung or pancreas. The proteins deduced from these sequences are identical to those obtained for the HT-29 line.

- 5) Cloning of the human cDNA into plasmid pCDNA3  
5 (Invitrogen V 790-20)

The PCR product obtained in 3) and also the plasmid are digested with the two restriction enzymes Kpn I and Xba I and then purified after migration on a 1% agarose gel using the GeneClean kit (Bio 101 reference 3105). After ligation with 100 ng of insert and 10 ng of vector and transformation (technique described in EXAMPLE I.3.g and i), the recombinant clones are verified by sequencing using the Applied Biosystems kit mentioned above.

15 EXAMPLE III

Cloning of mouse SR-p70 cDNA from AtT-20 (pituitary tumour) cells

1) Cell culturing of the line AtT-20

The cells are cultured in Ham F10 medium (GIBCO 31550-023) with the addition of 15% of horse serum (GIBCO 26050-047), 2.5% of foetal calf serum (GIBCO 10081-073) and 50 mg/l of gentamicin, to semi-confluence.

2) Preparation of the complementary DNA library

The library is produced as described in EXAMPLE I. 2 and 3 from the cells cultured above.

3) Screening of the library

a) Preparation of the membranes

The clones of the library are plated out on LB agar medium (Petri dishes 150 mm in diameter) coated with Biodyne A membranes (PALL reference BNNG 132). After one night at 37°C, the clones are transferred by contact onto fresh membranes. The latter are treated by depositing them on 3 mm Whatman paper soaked with the following solutions: 0.5 N NaOH, 1.5 M NaCl for 5 minutes, then

0.5 M Tris-HCl pH 8, 1.5 M NaCl for 5 minutes. After treatment with proteinase K in the following buffer: 10 mM Tris-HCl pH 8, 10 mM EDTA, 50 mM NaCl, 0.1% SDS, 100 µg/ml proteinase K, for one hour at room temperature, the membranes are washed copiously in 2 × SSC (sodium citrate, NaCl), dried and then incubated in an oven under vacuum at 80°C for 20 minutes.

b) Preparation of the probe

On the basis of monkey and human SR-p70 cDNA sequences, a first sequence was produced on a fragment amplified from line AtT-20 mRNA as described in EXAMPLE II.3 and 4, with the oligomers of the following compositions:

sense primer: GCC ATG CCT GTC TAC AAG  
antisense primer: ACC AGC TGG TTG ACG GAG.

On the basis of this sequence, an oligomeric probe specific for mouse was chosen and possesses the following composition:

GAG CAT GTG ACC GAC ATT G.

100 ng of the probe are labelled at the 3' end with 10 units of terminal transferase (Pharmacia) and 100 µCi of [ $\alpha$ -<sup>32</sup>P]dCTP 3000 Ci/mmol (Amersham reference PB 10205) in 10 µl of the following buffer: 30 mM Tris-HCl pH 7.6, 140 mM cacodylic acid, 1 mM CoCl<sub>2</sub>, 0.1 mM DTT for 15 minutes at 37°C. The radiolabelled nucleotides not incorporated are removed on a column of polyacrylamide P10 (Biorad, reference 1504144). The probe obtained has a specific activity of approximately 5 × 10<sup>8</sup> dpm/µg.

c) Prehybridization and hybridization

The membranes prepared in a) are prehybridized for 30 minutes at 42°C in 6 × SSC, 5 × Denhart's, 0.1% SDS, and then hybridized for a few hours in the same buffer with the addition of the probe prepared in b) in the proportion of 10<sup>6</sup> dpm/ml.

d) Washing and exposure of the membranes

The membranes are washed twice at room temperature in 2 × SSC/0.1% SDS buffer and then for one hour at 56°C in 6 × SSC/0.1% SDS. The hybridized clones are visualized with KODAK XOMAT films. A positive clone

containing the mouse SR-p70 is selected and hereinafter designated as SR-p70c.

4) Sequencing of mouse SR-p70 and analysis of the sequence

5       The sequence is obtained using the Applied Biosystem kit (reference 401628). The protein sequence deduced from mouse SR-p70c cDNA (Figure 7) exhibits a very strong homology with the human and monkey sequences, except in the N-terminal portion which diverges strongly (see Figure 9). Using the so-called PCR technique in a similar manner to that described in EXAMPLE II.3 and 4, a second 5' sequence (originating from the same AtT-20 library) was obtained (Figure 8). The deduced N-terminal protein sequence (sequence designated SR-p70a) is very similar to that deduced from human and monkey SR-p70 cDNAs (SR-p70a) (Figure 9). The line AtT-20 hence affords at least two SR-p70 transcripts. The latter 2 diverge in the N-terminal portion through different splicings.

EXAMPLE IV

- 20   1) Production of recombinant SR-p70 protein in *E. coli*  
a) Construction of the expression plasmid

      This consists in placing the COOH-terminal portion of the monkey SR-p70a protein, from the valine at position 427 to the COOH-terminal histidine at position 637, in fusion with the glutathione S-transferase (GST) of the plasmid vector pGEX-4T-3 (Pharmacia reference 27-4583). For this purpose, the corresponding insert of SR-p70a (position 1434 to 2066) was amplified by PCR with 10 ng of plasmid containing monkey SR-p70a cDNA. The  
30   nucleic acid primers are of the following composition:

sense primer: TTT GGA TCC GTC AAC CAG CTG GTG GGC CAG  
                  BamHI restriction site

antisense primer: AAA GTC GAC GTG GAT CTC GGC CTC C.  
                  Sal I site

The fragment obtained and also the vector are digested with the restriction enzymes BamHI and Sal I and cloning is carried out as described in EXAMPLE II.5. The selected clone is referred to as pG SR-p70.

- 5 b) Expression and purification of the GST-pSR-p70 fusion protein

This step was carried out using the "bulk GST purification module" kit (Pharmacia Reference 27-4570-01).

- 10 In outline, the recombinant clone was cultured at 37°C in one litre of 2 x YTA medium + 100 µg/ml ampicillin. At OD 0.8, expression is induced with 0.5 mM IPTG for 2 hours at 37°C. After centrifugation, the cell pellet is taken up in cold PBS and then sonicated by  
15 ultrasound. After the addition of 1% Triton X-100, the preparation is incubated for 30 minutes with agitation at room temperature. After centrifugation at 12,000 g for 10 minutes at 4°C, the supernatant is recovered. Purification is then carried out on a glutathione-  
20 Sepharose 4B affinity chromatography column. Binding and washing are carried out in PBS buffer and elution is carried out by competition with reduced glutathione. The final concentration is brought to 300 µg/ml of fusion protein.

- 25 2) Production of SR-p70a protein in COS-3 cells

- COS-3 cells are transfected with pSE1 plasmid DNA into which monkey SR-p70a cDNA has been cloned (EXAMPLE I.1), or with the vector pSE1 plasmid DNA as control, by the DEAE-dextran technique: the COS-3 cells are  
30 inoculated at  $5 \times 10^5$  cells per 6 cm dish in culture medium containing 5% of foetal bovine serum (EXAMPLE I.1). After culture, the cells are rinsed with PBS. 1 ml of the following mixture is added: medium containing 6.5 µg of DNA, 250 µg/ml of DEAE-dextran and 100 µM  
35 chloroquine. The cells are incubated at 37°C in 5% CO<sub>2</sub> for 4 to 5 hours. The medium is aspirated off, 2 ml of PBS containing 10% of DMSO are added and the cells are incubated for one minute, shaking the dishes gently. The

medium is aspirated off again and the cells are rinsed twice with PBS. The cells are then incubated at 37°C with medium containing 2% of foetal bovine serum for the period during which expression takes place, which is generally 3 days.

The SR-p70a protein is then analysed as described in EXAMPLE IV by immunoblotting.

#### EXAMPLE V

##### Preparation of specific antibodies

150 µg of proteins of the sample prepared according to EXAMPLE IV were used to immunize a rabbit (New Zealand male weighing 1.5 to 2 kg approximately). The immunizations were performed every 15 days according to the protocol described by Vaitukaitis, Methods in Enzymology, 1981, 73, 46. At the first injection, one volume of antigenic solution is emulsified with one volume of Freund's complete adjuvant (Sigma reference 4258). Five boosters were administered in Freund's incomplete adjuvant (Sigma reference 5506).

#### EXAMPLE VI

##### Detection of the SR-p70 protein: Western immunoblotting

##### 1) Materials used for immunoblotting

##### a) Cell lines used for immunoblotting

The following cell lines were cultured as described in the catalogue "Catalogue of cell lines and hybridomas, 7th edition, 1992" of the ATCC (American Type Culture Collection): COS-3, CV-1 (monkey kidney cell line), HT-29, U-373MG (human glioblastoma), MCF7 (human mammary adenocarcinoma), SKNAS (human neuroblastoma), cultured under the same conditions as COS-3), SK-N-MC (human neuroblastoma), IMR-32 (human neuroblastoma), CHP212 (human neuroblastoma cultured under the same conditions as CV-1), Sacs-2 (osteosarcoma), SK-OV-3 (ovarian adenocarcinoma) and SW 480 (human colon adenocarcinoma).

b) COS-3 cells transfected by SR-p70a cDNA

COS-3 cells were transfected as described in EXAMPLE IV.2. As a control, the cells were transfected with pSE1 plasmid DNA not containing recombinant SR-p70a cDNA.

2) Preparation of protein samples from a eukaryotic cell culture or from transfected cells

After culture, the cells are washed with PBS and then taken up in RIPA buffer (PBS with 1% NP40, 0.5% sodium deoxycholate, 0.5% SDS) supplemented with 10 µg/ml RNAse A, 20 µg/ml DNase 1, 2 µg/ml aprotinin, 0.5 µg/ml leupeptin, 0.7 µg/ml pepstatin and 170 µg/ml PMSF. The cells are sonicated by ultrasound at 4°C and left for 30 minutes at 4°C. After microcentrifugation at 12,000 rpm, the supernatant is recovered. The protein concentration is measured by the Bradford method.

3) Western blotting

5 or 50 µg of proteins (50 µg for the cell lines and 5 µg for transfected cells) are placed in 0.2 volume of the following 6 x electrophoresis buffer: 0.35 mM Tris-HCl pH 6.8, 10.3% SDS, 36% glycerol, 0.6 mM DTT, 0.012% bromophenol blue. The samples are applied and run in a 10% SDS-PAGE gel (30:0.8 Bis) and then electrotransferred onto a nitrocellulose membrane.

4) Visualization with the antibody

The membrane is incubated for 30 minutes in TBST blocking buffer (10 mM Tris-HCl pH 8, 150 mM NaCl, 0.2% Tween 20) with the addition of 5% of milk (GIBCO - SKIM MILK) at room temperature. The membrane is brought into contact successively with the anti-SR-p70 (αSR-p70) antibody in the same buffer for 16 hours at 4°C, washed 3 times for 10 minutes with TBST and then incubated for one hour at 37°C with a second, anti-rabbit immunoglobulin antibody coupled to peroxidase (SIGMA A055). After three washes of 15 minutes, the visualization is performed using the ECL kit (Amersham

RPN2106) by chemiluminescence.

In parallel, the same samples were subjected to visualization with an anti-p53 ( $\alpha$ p53) antibody (Sigma BP5312) followed by a second, anti-mouse immunoglobulin antibody.

## 5) Figures and results

### Figure 10: Immunoblot of the SR-p70 protein

Figure 10a: Detection of the recombinant SR-p70 protein

- columns 1 and 3: COS-3 transfected by the vector pSE1.
- columns 2 and 4: COS-3 transfected by plasmid pSE1 containing SR-p70a cDNA.
- columns 1 and 2: visualization with the anti-SR-p70 ( $\alpha$ SR-p70) antibody.
- columns 3 and 4: visualization with the anti-p53 ( $\alpha$ p53) antibody.

Figure 10b: Detection of the endogenous SR-p70 protein

- columns 1: COS-3; 2: CV-1; 3: HT-29; 4: U-373 MG; 5: MCF7; 6: SKNAS; 7: SK-N-MC; 8: IMR-32; 9: CHP212; 10: Saos-2; 11: SK-OV-3 and 12: SW480.

A: Visualization with the  $\alpha$ SR-p70 antibody

B: Visualization with the  $\alpha$ p53 antibody.

The  $\alpha$ SR-p70 antibody specifically recognizes the recombinant proteins (Figure 10a) and endogenous proteins (Figure 10b) and does not cross with p53. The analysis of human or monkey cell lines shows the SR-p70 protein, like p53, is generally weakly detectable. In contrast, when an accumulation of p53 exists, SR-p70 becomes, for its part also, more readily detectable (Figure 10b). A study by RT-PCR of the distribution of SR-p70 transcripts shows that the gene is expressed in all the cell types tested.

## EXAMPLE VII

Cloning of the SR-p70 gene and chromosomal localization

### 1) Cloning of SR-p70 gene

The library used is a cosmid library prepared

in the EXAMPLE III.3, with an SR-p70 DNA fragment labelled with  $^{32}\text{P}$  with the BRL "Random Primers DNA Labelling Systems" kit (reference 18187-013). The hybridization and washing buffers are treated by adding  
5 50% of formaldehyde. The last wash is carried out in 0.1 x SSC/0.1% SDS at 60°C. In a similar manner, the SR-p70 gene was isolated from a library prepared with C57 black mouse genomic DNA.

An analysis and a partial sequencing of the  
10 clones demonstrate the presence of 14 exons with a structure close to that of the p53 gene, in particular in the central portion where the size and positioning of the exons are highly conserved (Figure 12). This structure was partially defined in mouse and in man.

15 As an example, the human genomic sequences of the 3' region of intron 1, of exon 2, of intron 3 and of the 5' region of exon 3 are presented in Figure 13.

## 2) Chromosomal localization of the SR-p70 gene in man

This was carried out with human SR-70 gene DNA  
20 using the technique described by R. Slim et al., Hum. Genet., 1991, 88, 21-26. Fifty mitoses were analysed, more than 80% of which had double spots localized at 1p36 on both chromosomes and more especially at 1p36.2-1p36.3 (Figure 11). The identification of chromosome 1 and its  
25 orientation are based on the heterochromatin of the secondary constriction. The pictures were produced on a Zeiss Axiophot microscope, taken with a LHESA cooled CCD camera and treated with Optilab.

## EXAMPLE VIII

30 A) Demonstration of an mRNA coding for a deduced human SR-p70 protein possessing both a shorter N-terminal end and a divergence.

1) *Culturing of IMR-32 (human neuroblastoma) cells*

The cells were cultured as described in the catalogue "Catalogue of cell lines and hybridomas, 7th edition, 1992" of the ATCC (American Type Culture Collection).

2) *Preparation of the cDNA*

The RNA is prepared as described in Example I.2.a. The cDNA is prepared in a manner similar to that described in Example I.3, with 5 µg total RNA in a final volume of 20 µl using a poly(T)<sub>12</sub> primer and with cold nucleotides. The reaction is not interrupted with EDTA.

3) *Specific amplification of SR-p70 cDNA by the so-called PCR technique*

The polymerization is carried out with 2 µl of cDNA in 50 µl final with the buffer of the following composition: 50 mM Tris-HCl pH 9.2, 16 mM (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 1.75 mM MgCl<sub>2</sub>, in the presence of 10% DMSO, 0.4 mM NTP, 100 ng of each of the two nucleic acid primers and 3.5 units of the mixture of TAQ and PWO polymerases (Boehringer Mannheim, ref. 1681 842).

The primer pair is of the following composition:

sense primer: AGGCCGGCGTGGGGAAG (position 16 to 32, Figure 6)

antisense primer: CTTGGCGATCTGGCAGTAG (position 503 to 485, Figure 6).

The reaction is carried out for 30 cycles at 95°C/30 seconds, 58°C/1 minute and 68°C/2 minutes 30 seconds, followed by a final cycle of 68°C/10 minutes.

The PCR product is subjected to electrophoresis on a 1% agarose gel (TAE buffer). After ethidium bromide staining, two major bands are revealed: a band approximately 490 bp in size (expected size (see Figure 6)) and an additional band approximately 700 bp in size. The latter is extracted from the gel using the "Geneclean" kit (Bio 101, ref 1001 400). After a desalting on a column of polyacrylamide P10 (Biorad, ref

15011050), the fragment is subjected to a further PCR amplification for 10 cycles as described above.

4) Determination of the sequence of the amplified product

In a first step, the PCR product is removed from the oligonucleotides on a column of Sephacryl S-400 (Pharmacia 17-0609-01) and then desalted on a column of P10. The sequencing reaction is carried out using the Applied Biosystems kit (ref. 401 628) (373 DNA sequencer) with the antisense primer.

The sequence obtained is identical to the SR-p70 cDNA sequence (Example II.4) with an insertion of 198 bp between positions 217 and 218 (Figure 14). The deduced N-terminal protein sequence (sequence designated SR-p70d) is 49 amino acids shorter, with a divergence of the first 13 amino acids (sequence ID No. 13). There is hence coexistence of at least two different SR-p70 transcripts as already described for the mouse AtT-20 line.

B) Cloning of human SR-p70 and demonstration of an mRNA coding for a deduced human SR-p70 protein possessing the same N-terminal end as SR-p70d and a divergence in the C-terminal portion

1) Specific amplification of SR-p70 cDNA by the so-called PCR technique

The amplification was carried out as described in EXAMPLE VIII.A from purified RNA of IMR-32 cells with the primer pair of the following composition:

sense primer: GCG GCC ACG ACC GTG AC (position 160 to 176, sequence ID No. 11)

antisense primer: GGC AGC TTG GGT CTC TGG (position 1993 to 1976, Figure 6).

After removal of the excess primers on an S400 column and desalting on a P10 column, 1  $\mu$ l of the sample is subjected again to a PCR with the primer pair of the following composition:

sense primer: TAT CTC GAG CTG TAC GTC GGT GAC CCC  
XhoI (position

263 to 280, sequence ID No. 11)

antisense primer: ATA TCT AGA TCA GTG GAT CTC GGC CTC

XbaI

(position

1943 to 1926, Figure 6).

5    2) Cloning of the amplified product into plasmid pCDNA3

          The PCR product obtained in 1) is desalted on a P10 column, digested with the restriction enzymes XhoI and XbaI and then cloned into plasmid pCDNA3 as described in EXAMPLE II.5. Two recombinant clones are sequenced using the Applied Biosystems kit with the oligonucleotides specific for SR-p70 cDNA.

          The first sequence obtained corresponds to the complete sequence of the mRNA coding for SR-p70 described in EXAMPLE VIII.a. The deduced protein contains 587 amino acids (sequence ID No. 13 and Figure 16).

          The second sequence obtained is identical to the SR-p70d cDNA sequence described above, but with two deletions, of 149 bp and of 94 bp between positions 1049 and 1050 on the one hand, and between positions 1188 and 1189 on the other hand (sequence ID No. 14 and Figure 15). The protein sequence deduced from this second sequence reveals a protein having an N-terminal portion 49 amino acids shorter, with a divergence in the first 13 amino acids as well as a divergence of protein sequence between amino acids 350 and 397 (sequence ID No. 15 and Figure 16) (sequence designated SR-p70e). The deduced protein contains 506 amino acids.

C) Demonstration of an mRNA coding for a deduced human SR-p70 protein possessing a shorter N-terminal end

30    1) Culturing of SK-N-SH (human neuroblastoma) cells

          The cells are cultivated as described in the "Catalogue of cell lines and hybridomas, 7th edition, 1992" of the ATCC (American Type Culture Collection).

2) Preparation of the cDNA and amplification of SR-p70 cDNA by the so-called PCR technique

These steps are carried out as described in EXAMPLE VIII.A with the primer pair of the following composition:

sense primer: AGG GGA CGC AGC GAA ACC (position 128 to 145, Figure 17)

antisense primer: GGC AGC TTG GGT CTC TGG (position 1993 to 1976, Figure 6).

The sequencing is carried out with the Applied Biosystem kit with primers specific for SR-p70 cDNA, and reveals two cDNAs:

- a first cDNA corresponding to the mRNA coding for SR-p70a

- a second cDNA having a deletion of 98 bp between positions 24 and 25 (sequence ID No. 16 and Figure 15).

This deletion comprises the translation initiation ATG of SR-p70a. The protein deduced (designated SR-p70f) from this second cDNA possesses a translation initiation ATG downstream corresponding to an internal ATG of SR-p70a. The deduced protein hence contains 588 amino acids (sequence ID No. 17 and Figure 16) and is truncated with respect to the 48 N-terminal amino acids of SR-p70a.

D) Demonstration of an mRNA coding for human SR-p70b

1) Culturing of K562 cells

The cells are cultured as described in the "Catalogue of cell lines and hybridomas, 7th edition, 1992" of ATCC (American Type Culture Collection).

2) Preparation of the cDNA, amplification of SR-p70 cDNA by the so-called PCR technique and sequencing

These steps are carried out as described in EXAMPLE VIII.C.

The sequencing reveals two cDNAs:

A first cDNA corresponding to the mRNA coding for SR-p70a, and a second cDNA having a deletion of 94 bp

between positions 1516 and 1517 (sequence ID No. 18 and Figure 15). The deduced protein (designated SR-p70b) contains 199 amino acids and possesses a C-terminal sequence truncated by 137 amino acids relative to SR-p70a, with the last 4 amino acids divergent (sequence ID 5 No. 19 and Figure 21).

This cDNA is similar to the one described in EXAMPLE I relating to monkey SR-p70b.

The molecules described in this example (EXAMPLE 10 VIII.A, B, C and D) reveal SR-p70 variants which are the outcome of differential splicings of the primary mRNA, transcribed by the SR-p70 gene.

The SR-p70a is encoded by an mRNA composed of 14 exons (see EXAMPLE VII). This is the reference protein. 15 SR-p70b is the outcome of an insertion between exons 3 and 4 and of the absence of exons 11 and 13. SR-p70f is the outcome of the absence of exon 2. This example describes the existence of SR-p70 variants non-exhaustively, with a strong probability of existence of 20 other variants. Similarly, the existence of these variants described in this example, as well as SR-p70a, is not limited to the lines in which they have been demonstrated. In effect, studies performed by RT-PCR showed that these variants are to be found in the various 25 lines studied.

Furthermore, the initiation methionine of SR-p70f corresponds to an internal methionine of SR-p70a, suggesting the possibility of initiation downstream on the mRNA coding for SR-p70a.

#### 30 EXAMPLE IX

Obtaining a 5' sequence of human SR-p70a mRNA

1) Amplification of the 5' end of SR-p70 cDNA by PCR

The cell culturing and the preparations of total RNA and of cDNA are carried out as described in EXAMPLE 35 VIII.1 and 2. The RNA template is hydrolysed by incubation for 5 minutes at 65°C after the addition of 4 µl of 500 mM EDTA and 4 µl of 2 N NaOH. The sample is

then desalted on a P10 column. The cDNA is elongated at the 3' end with a dG tail as described in EXAMPLE I.3.d, in a final volume of 40  $\mu$ l. After the addition of 4  $\mu$ l of 500 mM EDTA and 4  $\mu$ l of 2 N NaOH, the cDNA is incubated at 65°C for 3 minutes and then desalted on a P10 column. PCR amplification is carried out as described in EXAMPLE VIII.3 with 8  $\mu$ l of cDNA and for 30 cycles with the primer pair of the following composition:

sense primer: C C C C C C C C C C C C C N (where N equals G, A or T)

antisense primer: CCATCAGCTCCAGGCTCTC (position 1167 to 1149, Figure 6).

After removal of the excess primers on an S-400 column and desalting on a P10 column, 1  $\mu$ l of the sample is subjected again to a PCR with the pair of the following composition:

sense primer: C C C C C C C C C C C C C N

antisense primer: CCAGGACAGGCGCAGATG (position 928 to 911, Figure 6).

The sample, passed again through an S-400 column and a P10 column, is subjected to a third amplification for 20 cycles with the following pair:

sense primer: C C C C C C C C C C C C C C N

antisense primer: CTTGGCGATCTGGCAGTAG (position 503 to 485, Figure 6).

## 2) Determination of the SR-p70 cDNA 5' sequence

The sequence is produced as described in EXAMPLE VIII.4. This sequence reveals a non-coding 5' region of at least 237 bases upstream of the initiation ATG of SR-p70a (Figure 17). By comparison of this sequence (obtained from the line IMR-32) with the one obtained from the line HT-29 in particular (Figure 6), two point differences (Figure 17: see bold characters) are revealed (G  $\rightarrow$  A and C  $\rightarrow$  T), positioned, respectively, at -20 and -30 from the initiation ATG of SR-p70a (Figures 6 and 17). This variability is located in exon 2 (Figure 13). It is not ruled out that this variability is also to be found within a coding frame as the outcome of an

alternative splicing as described in EXAMPLES III in mouse and VIII in man, or alternatively as the outcome of a translation initiation on a CTG (as has been demonstrated for FGFB (Proc. Natl. Acad. Sci USA, 1989, 86, 1836 - 1840)).

Similarly, it is not ruled out that this variability has a repercussion on the translation of SR-p70 or on the splicing of the primary RNA.

At all events, this variability, probably of allelic origin, may serve as a marker, either at genomic level (see EXAMPLE XI) or at mRNA level (see EXAMPLE X).

#### EXAMPLE X

1) Analysis by PCR of the transcriptional expression of SR-p70a in cell samples (RT-PCR)

Cell culturing (SK-N-AS, SK-N-MC, HT-29, U-373MG, SW480, IMR-32, CHP212) is carried out as described in Example VI.1.a (referred to the catalogue "Catalogue of cell lines and hybridomas, 7th edition 1992" of the ATCC).

The preparation of the cDNA and the PCR amplification are carried out as described in EXAMPLE VIII.2 and 3. The primer pair used is of the following composition:

sense primer: AGGGGACGCGAGCGAAACC (position 128 to 145, Figure 17)  
antisense primer: GGCAGCTTGGGTCTCTGG (position 1993 to 1976, Figure 6).

The samples are analysed by electrophoresis on a 1% agarose gel and visualization with ethidium bromide (Figure 18).

The size of the band obtained in the samples corresponds to the expected size (approximately 2 kb, Figures 6 and 17). The intensity of the bands obtained is reproducible. A reamplification of 1  $\mu$ l of the sample under the same conditions for 20 cycles reveals a band in each of the samples.

2) Determination of the sequence of the amplified products

After passage of the samples through S-400 and P10 columns, sequencing is carried out on an Applied Biosystems sequencer 373 with the reference kit 401 628. The primers used are, inter alia, the following:

		position	Figure
	AGGGGACGCAGCGAAACC	128 to 145	22
	CTTGCGGATCTGGCAGTAG	503 to 485	6
	GATGAGGTGGCTGGCTGGA	677 to 659	6
10	CCATCAGCTCCAGGCTCTC	1167 to 1149	6
	TGGTCAGGTTCTGCAGGTG	1605 to 1587	6
	GGCAGCTTGGGTCTCTGG	1993 to 1976	6

No protein difference in the SR-p70a was detected. However, sequences obtained reveal a double variability at positions -20 and -30 upstream of the initiation ATG of SR-p70a (Figures 6 and 17). This variability, probably of allelic origin, enables two classes of transcripts to be defined: a first class possessing a G at position -30 and a C at position -20 (class G<sup>-30</sup>/C<sup>-20</sup>) and a second class possessing a difference at two positions: an A at -30 and a T at -20 (class A<sup>-30</sup>/T<sup>-20</sup>).

First class: SK-N-AS, SK-N-MC, HT-29, U-373MG, SW480.

Second class: IMR-32, CHP212.

25 EXAMPLE XI

Analytical method of determination of the allelic distribution of the SR-p70 gene in a population of 10 persons

This allelic distribution is based on the allelic variability demonstrated in EXAMPLES IX and X:

- G<sup>-30</sup>/C<sup>-20</sup> allele possessing, respectively, a G and a C at positions -30 and -20 upstream of the initiation ATG of SR-p70a.
- A<sup>-30</sup>/T<sup>-20</sup> allele possessing, respectively, an A and

a T at the same positions.

This variability may be demonstrated by the use of restriction enzymes that differentiate the two alleles (Figure 13). As an example:

- 5     •    Enzyme Bpl I having a cleavage site only on the G<sup>-30</sup>/C<sup>-20</sup> allele in the zone of interest (this site encompasses both variable positions).
- Enzyme StyI having a cleavage site only on the A<sup>-30</sup>/T<sup>-20</sup> allele in the zone of interest.

10    1) Genomic amplification of exon 2 by PCR

The polymerization reaction is carried out with 500 ng of purified genomic DNA, in 50 µl final with the conditions described in Example VIII.3.

The primer pair is of the following position:

- 15    Sense primer:        CACCTACTCCAGGGATGC (position 1 to 18, Figure 13)  
     Antisense primer:   AGGAAATAGAACGTCGATGC (position 833 to 813, Figure 13).

The reaction is carried out for 30 cycles as described in EXAMPLE VIII.3.

- 20    After removal of the excess primer on an S-400 column and desalting on a P10 column, 1 µl of the sample is amplified again for 25 cycles under the same conditions with the following primer pair:

     Sense primer:        CAGGCCCACTTCGCTGCC (position 25 to 32, Figure 13)  
     Antisense primer:   CTGTCCCCAAGCTGATGAG (position 506 to 488, Figure 13).

- 25    The amplified products are subjected to electrophoresis on a 1% agarose gel (Figure 19-A).

2) Digestion with the restriction enzyme StyI

- 30    The samples are desalted beforehand on a P10 column and then digested with the restriction enzyme StyI (BRL 15442-015) in the buffer of the following composition: 50 mM Tris-HCl pH 8, 100 mM NaCl, 10 mM MgCl<sub>2</sub>, at 37°C for 30 min. The digestion products are analysed by electrophoresis on a 1% agarose gel (TAE buffer). Visualization is carried out by ethidium bromide

staining (Figure 19-B).

A band of 482 base pairs characterizes the G<sup>-30</sup>/C<sup>-20</sup> allele (Figures 13 and 19). The presence of a band of 376 base pairs and a band of 106 base pairs characterize the A<sup>-30</sup>/T<sup>-20</sup> allele (allele possessing a StyI cleavage site).

On the population of 10 persons, 2 persons exhibit the G<sup>-30</sup>/C<sup>-20</sup> and A<sup>-30</sup>/T<sup>-20</sup> alleles, the other 8 persons being homozygous with the G<sup>-30</sup>/C<sup>-20</sup> allele. The study of a fresh population of 9 persons demonstrated 3 heterozygous persons exhibiting the G<sup>-30</sup>/C<sup>-20</sup> and A<sup>-30</sup>/T<sup>-20</sup> alleles, the other 6 persons being homozygous for the G<sup>-30</sup>/C<sup>-20</sup> allele.

#### EXAMPLE XII

Test of reversion of transformation of the line SK-N-AS by transfection with SR-p70 cDNA

The expression vector used is described in EXAMPLE II.5 and shown diagrammatically in Figure 15. The method used is the so-called calcium phosphate method described by Graham et al. (Virology 1973, 54, 2, 536-539). The line is inoculated in the proportion of  $5 \times 10^5$  cells per dish 6 cm in diameter in 5 ml of the medium described in Example I.1. The cells are cultured at 37°C and with 5% CO<sub>2</sub> overnight. The transfection medium is prepared in the following manner: the following mixture is prepared by adding, in order, 1 ml of HEBS buffer (8 mg/ml NaCl, 370 µg/ml KCl, 125 µg/ml Na<sub>2</sub>HPO<sub>4</sub>·2H<sub>2</sub>O, 1 mg/ml dextrose, 5 mg/ml Hepes pH 7.05), 10 µg of the plasmid to be transfected and 50 µl of 2.5 M CaCl<sub>2</sub> added dropwise. The transfection medium is left for 30 min at room temperature and then added dropwise to the medium contained in the culture dish. The cells are incubated for 5 to 6 hours at 37°C/5% CO<sub>2</sub>. After the medium is aspirated off, 5 ml of fresh medium containing 2% of foetal bovine serum are added. After 48 hours at 37°C/5% CO<sub>2</sub>, the cells are rinsed with PBS, detached by trypsinization, diluted in 10 ml of culture medium (5%

foetal bovine serum) and plated out in a dish 10 cm in diameter (the dilution may be adjusted in accordance with the efficiency of transfection). After a further incubation for 10 hours (the time for the cells to adhere), the cells are subjected to selection by adding G418 at a final concentration of 600 µg/ml Geneticin equivalent for 15 to 21 days (the medium is changed every day). The clones obtained are then rinsed with PBS, fixed in 70% ethanol, dried, stained with 1% crystal violet and then counted.

Four plasmid transfections were carried out in duplicate:

- plasmid pCDNA3 without insert
- plasmid pCDNA3/SR-p70 containing human SR-p70a cDNA
- plasmid pCDNA3/SR-p70 Mut containing SR-p70a cDNA possessing a mutation at position 293 AA (R → H) which is analogous to the mutation 273 (R → H) in the DNA-binding domain of p53
- control without plasmid.

The result is expressed as the number of clones per dish.

	Experiment 1	Experiment 2	Mean
pCDNA3	172	353	262
pCDNA3/SR-p70	13	8	10
pCDNA3/SR-p70 Mut	92	87	89
Absence of plasmid	1	3	2

The number of clones obtained by transfection with plasmid pCDNA3/SR-p70 is 25-fold less than the number of clones obtained with the control pCDNA3 and 9-fold less than the number of clones obtained with pCDNA3/SR-p70 Mut, indicating a mortality or an arrest of cell division of the cells transfected with SR-p70 cDNA. This result is not the consequence of a toxicity in view of the clones obtained with the mutated SR-p70 cDNA, but probably of an apoptosis as has been demonstrated for the

p53 protein (Koshland et al., Sciences, 1993, 262, 1953-1981).

#### EXAMPLE XIII

##### Biological role of the SR-p70 protein

5           The structural homology between the DNA-binding  
domain of p53 and the central region of the SR-p70  
protein enables it to be inferred that SR-p70 is a  
transcription factor (see Figures 1 and 2). In effect,  
p53 (393 amino acids) consists of several functional  
10 domains. The N-terminal region (1-91 amino acids) is  
involved in the activation of transcription, and contains  
sites for interaction with different cellular and viral  
proteins. The central portion (amino acids 92 to 292)  
permits binding to the specific DNA sequences located in  
15 the promoter regions of certain genes (the majority of  
point mutations that inactivate p53 are localized in this  
region), and also possesses numerous sites for  
interaction with viral proteins which inhibit its  
activity. Finally, the last 100 amino acids of p53 are  
20 responsible for its oligomerization as well as for the  
regulation of the latter (Hainaut P., Current Opinion in  
Oncology, 1995, 7, 76-82; Prokocimer M., Blood, 1994, 84  
No. 8, 2391-2411).

25           The sequence homology between p53 and SR-p70 is  
significant, in particular as regards the amino acids  
involved directly in the interaction with DNA, suggesting  
that SR-p70 binds to the p53 sites on DNA. These amino  
acids correspond very exactly to what are referred to as  
the "hot spots", amino acids frequently mutated in human  
30 tumours (SWISS PROT: SW: P53\_human and Prokocimer M.,  
Blood, 1994, 84 No. 8, 2391-2411). From this homology, it  
may be deduced that the SR-p70 protein exerts a control  
over the activity of the genes regulated by p53, either  
independently of the latter or by forming heterooligomers  
35 with it.

Consequently, like p53, the products of the SR-  
p70 gene must be involved in the control and regulation

of the cell cycle, causing the cycle to stop (momentarily or permanently), and the implementation of programmes such as DNA repair, differentiation or cell death. The likelihood of the existence of "p53-like" activities had  
5 been strongly felt with the demonstration in p53<sup>-/-</sup> mice of activities of DNA repair and cell death in response to ionizing radiations (Strasser et al., Cell, 1994, 79, 329-339). The authors of the present invention have localized the human SR-p70 gene in the telomeric region  
10 of the short arm of chromosome 1, precisely at 1p36.2-36.3, the smallest deleted region (SRO) common to a majority of neuroblastomas and of other types of tumours (melanomas and sarcomas) (White et al., PNAS, 1995, 92, 5520-5524). This region of loss of heterozygosity (LOH)  
15 defines the locus of a tumour-suppressing gene whose loss of activity is considered to be the cause of tumour formation. It is important to recall that this region is also subject to "maternal imprinting"; the maternal allele is preferentially lost in neuroblastomas having  
20 the 1p36 deletion (without amplification of N-Myc) (Caron et al., Hum. Mol. Gen., 1995, 4, 535-539). The wide-type SR-p70 gene introduced into neuroblastoma cells and expressed therein permits the reversion of their trans-formation. The loss of this anti-oncogenic activity is  
25 hence associated with the development of the tumour. The 1p36 region possesses a syngeneic homology with the distal segment of the mouse chromosome 4. In this region, the curly tail (ct) gene (Beier et al., Mammalian Genome, 1995, 6, 269-272) involved in congenital malformations of  
30 the neural tube (NTM: *spina bifida*, *anencephaly*, etc). The ct mouse is the best animal model for studying these malformations. It is accepted that these malformations result from abnormalities of cell proliferation. Bearing in mind the nature of the SR-p70 gene and its chromosomal  
35 localization, one of the hypotheses is that SR-p70 could be the human homologue of ct and that, on this basis, the detection of early mutations and chromosomal abnormalities affecting this gene should permit, for example, as an application, the identification of persons

at risk (0.5-1% of newborn babies affected by NTM) and the implementation of preventive treatments (Neumann et al., Nature Genetics, 1994, 6, 357-362; Di Vinci et al., Int. J. Cancer, 1994, 59, 422-426; Moll et al., PNAS, 1995, 92, 4407-4411; Chen et al., Development, 1995, 121, 681-691).

#### EXAMPLE XIV

##### Allelic study of the SR-p70 gene

The GC and AT alleles are readily identified by StyI restriction of the PCR products of exon 2 (see Example XI). Hence it was possible to determine in this way, in GC/AT heterozygous individuals bearing neuroblastoma tumours, the lost SR-p70 allele (GC or AT), in spite of the presence of contaminating healthy tissue.

Surprisingly, when the same analysis is carried out on the RNA, a single allele is demonstrated independently of the presence or otherwise of a deletion and, still more surprisingly, in spite of the presence of healthy tissue. This suggests that the imprint (differential expression of the two alleles) would also exist in the contaminating tissue.

In order to verify this, the same analysis was repeated on the RNA originating from blood cells of healthy GC/AT heterozygous individuals. Only one of the two types of transcript was detected also in these cells. This result confirms the observation made on the tumour samples regarding the existence of a generalized genetic imprint for the SR-p70 gene.

The implications of this discovery are important, since it enables it to be postulated that a single sporadic mutation inactivating the active SR-p70 allele will give rise to a loss of activity, this potentially occurring in all the tissues.

The absence of precise data on the biological function of SR-p70 does not enable the consequences of this loss of SR-p70 activity for the cell to be measured.

Nevertheless, its strong homology with the p53 tumour-suppressing protein, as well as the demonstration that SR-p70 is a transcription factor capable of utilizing the P21<sup>waf</sup> promoter, suggests a role of this protein in the control of the cell cycle and in differentiation.

Knudson and Meadows, 1980 (New Eng. J. Med. 302: 1254-56), consider the IV-S neuroblastomas to be a collection of non-malignant cells from the neural crest carrying a mutation which interferes with their normal differentiation.

It is conceivable that the loss of SR-p70 activity, like the loss of p53 control over the cell cycle, favours the appearance of cellular abnormalities such as aneuploidy, amplification (described in the case of neuroblastomas) and other genetic reorganizations capable of causing cell transformation (Livingstone et al., 1992, Cell 71:923-25; Yin et al. 1992, Cell 72:937-48; Cross et al. 1995, Science 267:1353-56; Fukasawa et al. 1996, Science 271:1744-47). Neuroblastomas might hence arise originally from a temporary or permanent loss of activity of SR-p70, thereby favouring the occurrence of oncogenic events and hence tumour progression.

In the case of the 1p36 constitutional deletion described by Biegel et al., 1993 (Am. J. Hum. Genet. 52:176-82), IV-S neuroblastoma does indeed occur and the gene affected is NBS-1 (SR-p70).

In conclusion, what is described for neuroblastomas might also apply to other types of tumours, in particular those associated with reorganization of the end of the short arm of chromosome 1 (Report 2 international workshop on human chr 1 mapping 1995, Cytogenetics and Cell Genet. 72:113-154). From a therapeutic standpoint, the involvement of SR-p70 in the occurrence of tumours should lead to the avoidance of the use of mutagenic agents in chemotherapy, bearing in mind the risks of cell transformation by these products, and to the use, in preference to these products, of non-mutagenic substances which stimulate differentiation.

Moreover, the frequency of occurrence of the GC

and AT alleles is as follows: in the population,  $\text{Frequency(AT)}=0.15$ , and on a sample of 25 (neuroblastoma) patients,  $F(AT)=0.30$ . These statistics indicate that the AT allele could be a predisposing factor.

## SEQUENCE LISTING

## (1) GENERAL INFORMATION:

- (i) APPLICANT:  
 (A) NAME: sanofi  
 (B) STREET: 32-34 rue Marbeuf  
 (C) CITY: PARIS  
 (E) COUNTRY: FRANCE  
 (F) POSTAL CODE (ZIP): 75008  
 (G) TELEPHONE: 01 53 77 40 00  
 (H) TELEFAX: 01 53 77 41 33
- (ii) TITLE OF INVENTION: SR-p70
- (iii) NUMBER OF SEQUENCES: 40
- (iv) COMPUTER READABLE FORM:  
 (A) MEDIUM TYPE: Floppy disk  
 (B) COMPUTER: IBM PC Compatible  
 (C) OPERATING SYSTEM: PC-DOS/MS-DOS  
 (D) SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)

## (2) INFORMATION FOR SEQ ID NO: 1:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 2974 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA
- (vi) ORIGINAL SOURCE:  
 (A) ORGANISM: cebus apella
- (ix) FEATURE:  
 (A) NAME/KEY: CDS  
 (B) LOCATION: 156..2066

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

TGCCTCCCGG CCGCGGCACC CGCCCCGAGG CCTGTGCTCC TCGGAAGGGG ACGCAGCGAA	60
GCGGGGGCCC GCGCCAGGCC GGCCGGGACG GACGCCGATG CCGGAGCTG CGACGGCTGC	120
AGAGCGAGCT GCCCTCGGAG GCGGGTGTGA GGAAG ATG GCC CAG TCC ACC ACC	173
Met Ala Gln Ser Thr Thr	
1 5	
ACC TCC CCC GAT GGG GGC ACC ACG TTT GAG CAC CTC TGG AGC TCT CTG	221
Thr Ser Pro Asp Gly Gly Thr Thr Phe Glu His Leu Trp Ser Ser Leu	
10 15 20	
GAA CCA GAC AGC ACC TAC TTC GAC CTT CCC CAG TCA AGC CGG GGG AAT	269
Glu Pro Asp Ser Thr Tyr Phe Asp Leu Pro Gln Ser Ser Arg Gly Asn	
25 30 35	
AAT GAG GTG GTG GGT GGC ACG GAT TCC AGC ATG GAC GTC TTC CAC CTA	317
Asn Glu Val Val Gly Gly Thr Asp Ser Ser Met Asp Val Phe His Leu	
40 45 50	
GAG GGC ATG ACC ACA TCT GTC ATG GCC CAG TTC AAT TTG CTG AGC AGC	365
Glu Gly Met Thr Thr Ser Val Met Ala Gln Phe Asn Leu Leu Ser Ser	
55 60 65 70	
ACC ATG GAC CAG ATG AGC AGC CGC GCT GCC TCG GCC AGC CCG TAC ACC	413
Thr Met Asp Gln Met Ser Ser Arg Ala Ala Ser Ala Ser Pro Tyr Thr	
75 80 85	

CCG GAG CAC GCC AGC GTG CCC ACC GAT TCA CCC TAC GCA CAG CCC	461
Pro Glu His Ala Ala Ser Val Pro Thr His Ser Pro Tyr Ala Gln Pro	
90 95 100	
AGC TCC ACC TTC GAC ACC ATG TCG CCC GCG CCT GTC ATC CCC TCC AAC	509
Ser Ser Thr Phe Asp Thr Met Ser Pro Ala Pro Val Ile Pro Ser Asn	
105 110 115	
ACC GAC TAT CCC GGA CCC CAC CAC TTC GAG GTC ACT TTC CAG CAG TCC	557
Thr Asp Tyr Pro Gly Pro His His Phe Glu Val Thr Phe Gln Gln Ser	
120 125 130	
AGC ACG GCC AAG TCA GCC ACC TGG ACG TAC TCC CCA CTC TTG AAG AAA	605
Ser Thr Ala Lys Ser Ala Thr Trp Thr Tyr Ser Pro Leu Leu Lys Lys	
135 140 145 150	
CTC TAC TGC CAG ATC GCC AAG ACA TGC CCC ATC CAG ATC AAG GTG TCC	653
Leu Tyr Cys Gln Ile Ala Lys Thr Cys Pro Ile Gln Ile Lys Val Ser	
155 160 165	
GCC CCA CCG CCC CCG GGC ACC GCC ATC CGG GCC ATG CCT GTC TAC AAG	701
Ala Pro Pro Pro Pro Gly Thr Ala Ile Arg Ala Met Pro Val Tyr Lys	
170 175 180	
AAG GCG GAG CAC GTG ACC GAC ATC GTG AAG CCG TGC CCC AAC CAC GAG	749
Lys Ala Glu His Val Thr Asp Ile Val Lys Arg Cys Pro Asn His Glu	
185 190 195	
CTC GGG AGG GAC TTC AAC GAA GGA CAG TCT GCC CCA GCC AGC CAC CTC	797
Leu Gly Arg Asp Phe Asn Glu Gly Gln Ser Ala Pro Ala Ser His Leu	
200 205 210	
ATC CGT GTG GAA GGC AAT AAT CTC TCG CAG TAT GTG GAC GAC CCT GTC	845
Ile Arg Val Glu Gly Asn Asn Leu Ser Gln Tyr Val Asp Asp Pro Val	
215 220 225 230	
ACC GGC AGG CAG AGC GTC GTG GTG CCC TAT GAG CCA CCA CAG GTG GGG	993
Thr Gly Arg Gln Ser Val Val Val Pro Tyr Glu Pro Pro Gln Val Gly	
235 240 245	
ACA GAA TTC ACC ACC ATC CTG TAC AAC TTC ATG TGT AAC AGC AGC TGT	941
Thr Glu Phe Thr Thr Ile Leu Tyr Asn Phe Met Cys Asn Ser Ser Cys	
250 255 260	
GTG GGG GGC ATG AAC CGA CGG CCC CTC CTC ATC ATC ATC ACC CTG GAG	989
Val Gly Gly Met Asn Arg Arg Pro Ile Leu Ile Ile Ile Thr Leu Glu	
265 270 275	
ACG CGG GAT GGG CAG GTG CTG GGC CGC CGG TCC TTC GAG GGC CGC ATC	1037
Thr Arg Asp Gly Gln Val Leu Gly Arg Arg Ser Phe Glu Gly Arg Ile	
280 285 290	
TGC GCC TGT CCT GGC CGC GAC CGA AAA GCC GAT GAG GAC CAC TAC CGG	1085
Cys Ala Cys Pro Gly Arg Asp Arg Lys Ala Asp Glu Asp His Tyr Arg	
295 300 305 310	
GAG CAG CAG GCC TTG AAT GAG AGC TCC GCC AAG AAC GGG GCT GCC AGC	1133
Glu Gln Gln Ala Leu Asn Glu Ser Ser Ala Lys Asn Gly Ala Ala Ser	
315 320 325	
AAG CGC GCC TTC AAG CAG AGT CCC CCT GCC GTC CCC GCC CTG GGC CCG	1181
Lys Arg Ala Phe Lys Gln Ser Pro Pro Ala Val Pro Ala Leu Gly Pro	
330 335 340	
GGT GTG AAG AAG CGG CGG CAC GGA GAC GAG GAC ACG TAC TAC CTG CAG	1229
Gly Val Lys Lys Arg Arg His Gly Asp Glu Asp Thr Tyr Tyr Leu Gln	
345 350 355	
GTG CGA GGC CGC GAG AAC TTC GAG ATC CTG ATG AAG CTG AAG GAG AGC	1277
Val Arg Gly Arg Glu Asn Phe Glu Ile Leu Met Lys Leu Lys Glu Ser	
360 365 370	

CTG GAG CTG ATG GAG TTG GTG CCG CAG CCG CTG GTA GAC TCC TAT CGG Leu Glu Leu Met Glu Leu Val Pro Gln Pro Leu Val Asp Ser Tyr Arg 375 380 385 390	1325
CAG CAG CAG CAG CTC CTA CAG AGG CCG AGT CAC CTA CAG CCC CCA TCC Gln Gln Gln Gln Leu Leu Gln Arg Pro Ser His Leu Gln Pro Pro Ser 395 400 405	1373
TAC GGG CCG GTC CTC TCG CCC ATG AAC AAG GTG CAC GGG GGC GTG AAC Tyr Gly Pro Val Leu Ser Pro Met Asn Lys Val His Gly Val Asn 410 415 420	1421
AAG CTG CCC TCC GTC AAC CAG CTG GTG GGC CAG CCT CCC CCG CAC AGC Lys Leu Pro Ser Val Asn Gln Leu Val Gly Gln Pro Pro Pro His Ser 425 430 435	1469
TCG GCA GCT ACA CCC AAC CTG GGA CCT GTG GGC TCT GGG ATG CTC AAC Ser Ala Ala Thr Pro Asn Leu Gly Pro Val Gly Ser Gly Met Leu Asn 440 445 450	1517
AAC CAC GGC CAC GCA GTG CCA GCC AAC AGC GAG ATG ACC AGC AGC CAC Asn His Gly His Ala Val Pro Ala Asn Ser Glu Met Thr Ser Ser His 455 460 465 470	1565
GGC ACC CAG TCC ATG GTC TCG GGG TCC CAC TGC ACT CCG CCA CCC CCC Gly Thr Gln Ser Met Val Ser Gly Ser His Cys Thr Pro Pro Pro Pro 475 480 485	1613
TAC CAC GCC GAC CCC AGC CTC GTC AGT TTT TTA ACA GGA TTG GGG TGT Tyr His Ala Asp Pro Ser Leu Val Ser Phe Leu Thr Gly Leu Gly Cys 490 495 500	1661
CCA AAC TGC ATC GAG TAT TTC ACG TCC CAG GGG TTA CAG AGC ATT TAC Pro Asn Cys Ile Glu Tyr Phe Thr Ser Gln Gly Leu Gln Ser Ile Tyr 505 510 515	1709
CAC CTG CAG AAC CTG ACC ATC GAG GAC CTG GGG GCC CTG AAG ATC CCC His Leu Gln Asn Leu Thr Ile Glu Asp Leu Gly Ala Leu Lys Ile Pro 520 525 530	1757
GAG CAG TAT CGC ATG ACC ATC TGG CCG GGC CTG CAG GAC CTG AAG CAG Glu Gln Tyr Arg Met Thr Ile Trp Arg Gly Leu Gln Asp Leu Lys Gln 535 540 545 550	1805
GGC CAC GAC TAC GGC GCC GCC GCG CAG CAG CTG CTC CGC TCC AGC AAC Gly His Asp Tyr Gly Ala Ala Ala Gln Gln Leu Leu Arg Ser Ser Asn 555 560 565	1853
GCG GCC GCC ATT TCC ATC GGC GGC TCC GGG GAG CTG CAG CGC CAG CGG Ala Ala Ala Ile Ser Ile Gly Gly Ser Gly Glu Leu Gln Arg Gln Arg 570 575 580	1901
GTC ATG GAG GCC GTG CAC TTC CGC GTG CGC CAC ACC ATC ACC ATC CCC Val Met Glu Ala Val His Phe Arg Val Arg His Thr Ile Thr Ile Pro 585 590 595	1949
AAC CGC GGC GGC CCC GGC GCC GGC CCC GAC GAG TGG GCG GAC TTC GGC Asn Arg Gly Gly Pro Gly Ala Gly Pro Asp Glu Trp Ala Asp Phe Gly 600 605 610	1997
TTC GAC CTG CCC GAC TGC AAG GCC CCG AAG CAG CCC ATC AAG GAG GAG Phe Asp Leu Pro Asp Cys Lys Ala Arg Lys Pro Ile Lys Glu Glu 615 620 625 630	2045
TTC ACG GAG GCC GAG ATC CAC TGAGGGGCCG GGCCAGCCCA GAGCCTGTGC Phe Thr Glu Ala Glu Ile His 635	2096
CACCGCCCGAG AGACCCAGCG CGCCTCGCTC TCCTTCCTGT GTCCAAAACG GCCTCCGGAG	2156
GCAGGGCCCTC CAGGCTGTGC CCGGGGSAAG GCAAGGTCCG GCCCATGCC CGGCACCTCA	2216

```

CGGGCCCCAG GAGAGGCCCA GCCACCAAAG CCGCCTGCGG ACAGCTGTAG TCACCTGCAG 2276
AACCTTCTGG AGCTGCCCTA ATGCTGGGCT TCGGGGGCAG GGGCCGGCCC ACTCTCAGCC 2336
CTGCCACTGC CGGCGTGTCT CCATGGCAGG CGTGGGTGGG GACCGCAGTG TCAGCTCCGA 2396
CCTCCAGGCC TCATCTAGA GACTCTGTCA TCTGCCGATC AAGCAAGGTC CTTCAGAGG 2456
AAAGAATCCT CTTCGCTGGT GGACTGCCAA AAAGTATTTT GCGACATCTT TTGTTCTGG 2516
AGATGGTGTA GCAGCCCAAGC GACTGTGTCT GAAACACCGT GCATTTTCAG GGAATGTCCC 2576
TAACGGGCTG GGGACTCTCT CTGCTGGACT TGGGAGTGGC CTTTGCCCCC AGCACACTGT 2636
ATTCTGGGGG ACCGCCCTCT TCCTGCCCTT AACAAACCAC AAAGTGTGTC TGAATTTGGA 2696
GAAACTGGGG GAAGCGGCAG CCCCTCCCAG GTGCGGGAAG CATCTGGTAC CGCTCGGCC 2756
AGTGGCCCTC AGCTGGCCA CAGTCACCTC TCCTTGGGGA ACCCTGGGGA GAAAGGGACA 2816
GCCTGTCTCT AGAGACCGG AAATTGTCAA TATTGTATAA AATGATACCC TTTTCTAC 2874

```

## (2) INFORMATION FOR SEQ ID NO: 2:

- (1) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 637 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear

## (11) MOLECULE TYPE: protein

## (11) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

```

Met Ala Gln Ser Thr Thr Thr Ser Pro Asp Gly Gly Thr Thr Phe Glu
 1           5           10           15
His Leu Trp Ser Ser Leu Glu Pro Asp Ser Thr Tyr Phe Asp Leu Pro
          20           25           30
Gln Ser Ser Arg Gly Asn Asn Glu Val Val Gly Gly Thr Asp Ser Ser
          35           40           45
Met Asp Val Phe His Leu Glu Gly Met Thr Thr Ser Val Met Ala Gln
          50           55           60
Phe Asn Leu Leu Ser Ser Thr Met Asp Gln Met Ser Ser Arg Ala Ala
          65           70           75           80
Ser Ala Ser Pro Tyr Thr Thr Pro Glu His Ala Ser Val Pro Thr His
          85           90           95
Ser Pro Tyr Ala Gln Pro Ser Ser Thr Phe Asp Thr Met Ser Pro Ala
          100          105          110
Pro Val Ile Pro Ser Asn Thr Asp Tyr Pro Gly Pro His His Phe Glu
          115          120          125
Val Thr Phe Gln Gln Ser Ser Thr Ala Lys Ser Ala Thr Trp Thr Tyr
          130          135          140
Ser Pro Leu Leu Lys Lys Leu Tyr Cys Gln Ile Ala Lys Thr Cys Pro
          145          150          155          160
Ile Gln Ile Lys Val Ser Ala Pro Pro Pro Gly Thr Ala Ile Arg
          165          170          175
Ala Met Pro Val Tyr Lys Lys Ala Glu His Val Thr Asp Ile Val Lys
          180          185          190
Arg Cys Pro Asn His Glu Leu Gly Arg Asp Phe Asn Glu Gly Gln Ser

```

195	200	205
Ala Pro Ala Ser His Leu Ile Arg Val Glu Gly Asn Asn Leu Ser Gln 210 215 220		
Tyr Val Asp Asp Pro Val Thr Gly Arg Gln Ser Val Val Val Pro Tyr 225 230 235 240		
Glu Pro Pro Gln Val Gly Thr Glu Phe Thr Thr Ile Leu Tyr Asn Phe 245 250 255		
Met Cys Asn Ser Ser Cys Val Gly Gly Met Asn Arg Arg Pro Ile Leu 260 265 270		
Ile Ile Ile Thr Leu Glu Thr Arg Asp Gly Gln Val Leu Gly Arg Arg 275 280 285		
Ser Phe Glu Gly Arg Ile Cys Ala Cys Pro Gly Arg Asp Arg Lys Ala 290 295 300		
Asp Glu Asp His Tyr Arg Glu Gln Gln Ala Leu Asn Glu Ser Ser Ala 305 310 315 320		
Lys Asn Gly Ala Ala Ser Lys Arg Ala Phe Lys Gln Ser Pro Pro Ala 325 330 335		
Val Pro Ala Leu Gly Pro Gly Val Lys Lys Arg Arg His Gly Asp Glu 340 345 350		
Asp Thr Tyr Tyr Leu Gln Val Arg Gly Arg Glu Asn Phe Glu Ile Leu 355 360 365		
Met Lys Leu Lys Glu Ser Leu Glu Leu Met Glu Leu Val Pro Gln Pro 370 375 380		
Leu Val Asp Ser Tyr Arg Gln Gln Gln Gln Leu Leu Gln Arg Pro Ser 385 390 395 400		
His Leu Gln Pro Pro Ser Tyr Gly Pro Val Leu Ser Pro Met Asn Lys 405 410 415		
Val His Gly Gly Val Asn Lys Leu Pro Ser Val Asn Gln Leu Val Gly 420 425 430		
Gln Pro Pro Pro His Ser Ser Ala Ala Thr Pro Asn Leu Gly Pro Val 435 440 445		
Gly Ser Gly Met Leu Asn Asn His Gly His Ala Val Pro Ala Asn Ser 450 455 460		
Glu Met Thr Ser Ser His Gly Thr Gln Ser Met Val Ser Gly Ser His 465 470 475 480		
Cys Thr Pro Pro Pro Pro Tyr His Ala Asp Pro Ser Leu Val Ser Phe 485 490 495		
Leu Thr Gly Leu Gly Cys Pro Asn Cys Ile Glu Tyr Phe Thr Ser Gln 500 505 510		
Gly Leu Gln Ser Ile Tyr His Leu Gln Asn Leu Thr Ile Glu Asp Leu 515 520 525		
Gly Ala Leu Lys Ile Pro Glu Gln Tyr Arg Met Thr Ile Trp Arg Gly 530 535 540		
Leu Gln Asp Leu Lys Gln Gly His Asp Tyr Gly Ala Ala Ala Gln Gln 545 550 555 560		
Leu Leu Arg Ser Ser Asn Ala Ala Ala Ile Ser Ile Gly Gly Ser Gly 565 570 575		
Glu Leu Gln Arg Gln Arg Val Met Glu Ala Val His Phe Arg Val Arg		

580	595	600	605
His Thr Ile Thr Ile Pro Asn Arg Gly Gly Pro Gly Ala Gly Pro Asp			
595	600	605	
Glu Trp Ala Asp Phe Gly Phe Asp Leu Pro Asp Cys Lys Ala Arg Lys			
610	615	620	
Gln Pro Ile Lys Glu Glu Phe Thr Glu Ala Glu Ile His			
625	630	635	

## (2) INFORMATION FOR SEQ ID NO: 3:

- (1) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 2034 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(11) MOLECULE TYPE: cDNA

(VI) ORIGINAL SOURCE:  
 (A) ORGANISM: Cebus apella

(IX) FEATURE:  
 (A) NAME/KEY: CDS  
 (B) LOCATION: 156..1652

## (X1) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

TCGCTCCCGG	CCGCGCGCACC	CGCCCGGAGG	CCTGTGCTCC	TGCGAAGGGG	ACGCAGCGAA	60
GCGCGGGGCC	CGCGCAGGCC	GGCCGGGACG	GACGCCGATG	CCCGGAGGTG	CGACGGCTGC	120
AGAGCGAGCT	GCCTCTGGAG	GCCCGGTGTGA	GGAAG	ATG GCC CAG TCC ACC ACC		173
				Met Ala Gln Ser Thr Thr		
				1 5		
ACC TCC CCC GAT GGG GGC ACC ACG TTT GAG CAC CTC TGG AGC TCT CTG						221
Thr Ser Pro Asp Gly Gly Thr Thr Phe Glu His Leu Trp Ser Ser Leu						
10 15 20						
GAA CCA GAC AGC ACC TAC TTC GAC CTT CCC CAG TCA AGC CGG GGG AAT						269
Glu Pro Asp Ser Thr Tyr Phe Asp Leu Pro Gln Ser Ser Arg Gly Asn						
25 30 35						
AAT GAG GTG GTG GGT GGC ACG GAT TCC AGC ATG GAC GTC TTC CAC CTA						317
Asn Glu Val Val Gly Gly Thr Asp Ser Ser Met Asp Val Phe His Leu						
40 45 50						
GAG GGC ATG ACC ACA TCT GTC ATG GCC CAG TTC AAT TTG CTG AGC AGC						365
Glu Gly Met Thr Thr Ser Val Met Ala Gln Phe Asn Leu Leu Ser Ser						
55 60 65 70						
ACC ATG GAC CAG ATG AGC AGC CGC GCT GCC TCG GCC AGC CCG TAC ACC						413
Thr Met Asp Gln Met Ser Ser Arg Ala Ala Ser Ala Ser Pro Tyr Thr						
75 80 85						
CCG GAG CAC GCC GCC AGC GTG CCC ACC CAT TCA CCC TAC GCA CAG CCC						461
Pro Glu His Ala Ala Ser Val Pro Thr His Ser Pro Tyr Ala Gln Pro						
90 95 100						
AGC TCC ACC TTC GAC ACC ATG TCG CCC GCG CCT GTC ATC CCC TCC AAC						509
Ser Ser Thr Phe Asp Thr Met Ser Pro Ala Pro Val Ile Pro Ser Asn						
105 110 115						
ACC GAC TAT CCC GGA CCC CAC CAC TTC GAG GTC ACT TTC CAG CAG TCC						557
Thr Asp Tyr Pro Gly Pro His His Phe Glu Val Thr Phe Gln Gln Ser						
120 125 130						
AGC ACG GCC AAG TCA GCC ACC TGG ACG TAC TCC CCA CTC TTG AAG AAA						605

Ser	Thr	Ala	Lys	Ser	Ala	Thr	Trp	Thr	Tyr	Ser	Pro	Leu	Leu	Lys	Lys	
135					140					145					150	
CTC	TAC	TGC	CAG	ATC	GCC	AAG	ACA	TGC	CCC	ATC	CAG	ATC	AAG	GTG	TCC	653
Leu	Tyr	Cys	Gln	Ile	Ala	Lys	Thr	Cys	Pro	Ile	Gln	Ile	Lys	Val	Ser	
			155						160					165		
GCC	CCA	CCG	CCC	CCG	GGC	ACC	GCC	ATC	CGG	GCC	ATG	CCT	GTC	TAC	AAG	701
Ala	Pro	Pro	Pro	Pro	Gly	Thr	Ala	Ile	Arg	Ala	Met	Pro	Val	Tyr	Lys	
			170					175					180			
AAG	GCG	GAG	CAC	GTG	ACC	GAC	ATC	GTG	AAG	CGC	TGC	CCC	AAC	CAC	GAG	749
Lys	Ala	Glu	His	Val	Thr	Asp	Ile	Val	Lys	Arg	Cys	Pro	Asn	His	Glu	
		185					190					195				
CTC	GGG	AGG	GAC	TTC	AAC	GAA	GGA	CAG	TCT	GCC	CCA	GCC	AGC	CAC	CTC	797
Leu	Gly	Arg	Asp	Phe	Asn	Glu	Gly	Gln	Ser	Ala	Pro	Ala	Ser	His	Leu	
	200					205					210					
ATC	CGT	GTG	GAA	GGC	AAT	AAT	CTC	TCG	CAG	TAT	GTG	GAC	GAC	CCT	GTG	845
Ile	Arg	Val	Glu	Gly	Asn	Asn	Leu	Ser	Gln	Tyr	Val	Asp	Asp	Pro	Val	
	215				220					225					230	
ACC	GGC	AGG	CAG	AGC	GTG	GTG	GTG	CCC	TAT	GAG	CCA	CCA	CAG	GTG	GGG	893
Thr	Gly	Arg	Gln	Ser	Val	Val	Val	Pro	Tyr	Glu	Pro	Pro	Gln	Val	Gly	
				235					240					245		
ACA	GAA	TTC	ACC	ACC	ATC	CTG	TAC	AAC	TTC	ATG	TGT	AAC	AGC	AGC	TGT	941
Thr	Glu	Phe	Thr	Thr	Ile	Leu	Tyr	Asn	Phe	Met	Cys	Asn	Ser	Ser	Cys	
		250						255					260			
GTG	GGG	GGC	ATG	AAC	CGA	CGG	CCC	ATC	CTC	ATC	ATC	ATC	ACC	CTG	GAG	989
Val	Gly	Gly	Met	Asn	Arg	Arg	Pro	Ile	Leu	Ile	Ile	Ile	Thr	Leu	Glu	
	265						270					275				
ACG	CGG	GAT	GGG	CAG	GTG	CTG	GGC	CGC	CGG	TCC	TTC	GAG	GGC	CGC	ATC	1037
Thr	Arg	Asp	Gly	Gln	Val	Leu	Gly	Arg	Arg	Ser	Phe	Glu	Gly	Arg	Ile	
	280					285				290						
TGC	GCC	TGT	CCT	GGC	CGC	GAC	CGA	AAA	GCC	GAT	GAG	GAC	CAC	TAC	CGG	1085
Cys	Ala	Cys	Pro	Gly	Arg	Asp	Arg	Lys	Ala	Asp	Glu	Asp	His	Tyr	Arg	
	295				300				305					310		
GAG	CAG	CAG	GCC	TTG	AAT	GAG	AGC	TCC	GCC	AAG	AAC	GGG	GCT	GCC	AGC	1133
Glu	Gln	Gln	Ala	Leu	Asn	Glu	Ser	Ser	Ala	Lys	Asn	Gly	Ala	Ala	Ser	
			315						320				325			
AAG	CGC	GCC	TTC	AAG	CAG	AGT	CCC	CCT	GCC	GTC	CCC	GCC	CTG	GGC	CCG	1181
Lys	Arg	Ala	Phe	Lys	Gln	Ser	Pro	Pro	Ala	Val	Pro	Ala	Leu	Gly	Pro	
		330						335					340			
GGT	GTG	AAG	AAG	CGG	CGG	CAC	GGA	GAC	GAG	GAC	ACG	TAC	TAC	CTG	CAG	1229
Gly	Val	Lys	Lys	Arg	Arg	His	Gly	Asp	Glu	Asp	Thr	Tyr	Tyr	Leu	Gln	
	345					350						355				
GTG	CGA	GGC	CGC	GAG	AAC	TTC	GAG	ATC	CTG	ATG	AAG	CTG	AAG	GAG	AGC	1277
Val	Arg	Gly	Arg	Glu	Asn	Phe	Glu	Ile	Leu	Met	Lys	Leu	Lys	Glu	Ser	
	360				365						370					
CTG	GAG	CTG	ATG	GAG	TTG	GTG	CCG	CAG	CCG	CTG	GTA	GAC	TCC	TAT	CGG	1325
Leu	Glu	Leu	Met	Glu	Leu	Val	Pro	Gln	Pro	Leu	Val	Asp	Ser	Tyr	Arg	
	375				380				385						390	
CAG	CAG	CAG	CAG	CTC	CTA	CAG	AGG	CCG	AGT	CAC	CTA	CAG	CCC	CCA	TCC	1373
Gln	Gln	Gln	Gln	Leu	Leu	Gln	Arg	Pro	Ser	His	Leu	Gln	Pro	Pro	Ser	
				395					400					405		
TAC	GGG	CCG	GTC	CTC	TCG	CCC	ATG	AAC	AAG	GTG	CAC	GGG	GGC	GTG	AAC	1421
Tyr	Gly	Pro	Val	Leu	Ser	Pro	Met	Asn	Lys	Val	His	Gly	Gly	Val	Asn	
		410					415					420				
AAG	CTG	CCC	TCC	GTC	AAC	CAG	CTG	GTG	GGC	CAG	CCT	CCC	CCG	CAC	AGC	1469

Lys Leu Pro Ser Val Asn Gln Leu Val Gly Gln Pro Pro Pro His Ser  
425 430 435 1517

TCG GCA GGT ACA CCC AAC CTG GGA CCT GTG GGC TCT GGG ATG CTC AAC  
Ser Ala Ala Thr Pro Asn Leu Gly Pro Val Gly Ser Gly Met Leu Asn  
440 445 450

AAC CAC GGC CAC GCA GTG CCA GCC AAC AGC GAG ATG ACC AGC AGC CAC  
Asn His Gly His Ala Val Pro Ala Asn Ser Glu Met Thr Ser Ser His  
455 460 465 470 1565

GGC ACC CAG TCC ATG GTC TCG GGG TCC CAC TGC ACT CCG CCA CCC CCC  
Gly Thr Gln Ser Met Val Ser Gly Ser His Cys Thr Pro Pro Pro Pro  
475 480 485 1613

TAC CAC GCC GAC CCC AGC CTC GTC AGG ACC TGG GGG CCC TGAAGATCCC  
Tyr His Ala Asp Pro Ser Leu Val Arg Thr Trp Gly Pro 1662

CGAGCAGTAT CGCATGACCA TCTGGCGGGG CTGCAGGAC CTGAAGCAGG GCCACGACTA 1722

CGGCGCGCGCC GCGCAGCAGC TGCTCCGCTC CAGCAACGGG GCGGCCATTTC CCATCGGCGG 1782

CTCCGGGGAG CTGCAGCGCC AGCGGGTCAT GGAGGCCGTG CACTTCCGCG TGGCCACAC 1842

CATCACCATC CCCAACCGCG GCGGCCCCGG CCGCGGCCCG GACGAGTGGG CGGACTTCGG 1902

CTTCGACCTG CCCGACTGCA AGGCCCGCAA GCAGCCCATC AAGGAGGAGT TCACGGAGGC 1962

CGAGATCCAC TGAGGGGGCG GGGCCAGCCA GAGCCTGTGC CACCGCCCGAG AGACCCAGGC 2022

CGCCTCGCTC TC 2034

## (2) INFORMATION FOR SEQ ID NO: 4:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 499 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(11) MOLECULE TYPE: protein

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

Met Ala Gln Ser Thr Thr Thr Ser Pro Asp Gly Gly Thr Thr Phe Glu  
1 5 10 15

His Leu Trp Ser Ser Leu Glu Pro Asp Ser Thr Tyr Phe Asp Leu Pro  
20 25 30

Gln Ser Ser Arg Gly Asn Asn Glu Val Val Gly Gly Thr Asp Ser Ser  
35 40 45

Met Asp Val Phe His Leu Glu Gly Met Thr Thr Ser Val Met Ala Gln  
50 55 60

Phe Asn Leu Leu Ser Ser Thr Met Asp Gln Met Ser Ser Arg Ala Ala  
65 70 75 80

Ser Ala Ser Pro Tyr Thr Pro Glu His Ala Ala Ser Val Pro Thr His  
85 90 95

Ser Pro Tyr Ala Gln Pro Ser Ser Thr Phe Asp Thr Met Ser Pro Ala  
100 105 110

Pro Val Ile Pro Ser Asn Thr Asp Tyr Pro Gly Pro His His Phe Glu  
115 120 125

Val Thr Phe Gln Gln Ser Ser Thr Ala Lys Ser Ala Thr Trp Thr Tyr  
130 135 140

Ser Pro Leu Leu Lys Lys Leu Tyr Cys Gln Ile Ala Lys Thr Cys Pro  
 145 150 160  
 Ile Gln Ile Lys Val Ser Ala Pro Pro Pro Gly Thr Ala Ile Arg  
 165 170 175  
 Ala Met Pro Val Tyr Lys Lys Ala Glu His Val Thr Asp Ile Val Lys  
 180 185 190  
 Arg Cys Pro Asn His Glu Leu Gly Arg Asp Phe Asn Glu Gly Gln Ser  
 195 200 205  
 Ala Pro Ala Ser His Leu Ile Arg Val Glu Gly Asn Asn Leu Ser Gln  
 210 215 220  
 Tyr Val Asp Asp Pro Val Thr Gly Arg Gln Ser Val Val Val Pro Tyr  
 225 230 235 240  
 Glu Pro Pro Gln Val Gly Thr Glu Phe Thr Thr Ile Leu Tyr Asn Phe  
 245 250 255  
 Met Cys Asn Ser Ser Cys Val Gly Gly Met Asn Arg Arg Pro Ile Leu  
 260 265 270  
 Ile Ile Ile Thr Leu Glu Thr Arg Asp Gly Gln Val Leu Gly Arg Arg  
 275 280 285  
 Ser Phe Glu Gly Arg Ile Cys Ala Cys Pro Gly Arg Asp Arg Lys Ala  
 290 295 300  
 Asp Glu Asp His Tyr Arg Glu Gln Gln Ala Leu Asn Glu Ser Ser Ala  
 305 310 315 320  
 Lys Asn Gly Ala Ala Ser Lys Arg Ala Phe Lys Gln Ser Pro Pro Ala  
 325 330 335  
 Val Pro Ala Leu Gly Pro Gly Val Lys Lys Arg Arg His Gly Asp Glu  
 340 345 350  
 Asp Thr Tyr Tyr Leu Gln Val Arg Gly Arg Glu Asn Phe Glu Ile Leu  
 355 360 365  
 Met Lys Leu Lys Glu Ser Leu Glu Leu Met Glu Leu Val Pro Gln Pro  
 370 375 380  
 Leu Val Asp Ser Tyr Arg Gln Gln Gln Gln Leu Leu Gln Arg Pro Ser  
 385 390 395 400  
 His Leu Gln Pro Pro Ser Tyr Gly Pro Val Leu Ser Pro Met Asn Lys  
 405 410 415  
 Val His Gly Gly Val Asn Lys Leu Pro Ser Val Asn Gln Leu Val Gly  
 420 425 430  
 Gln Pro Pro Pro His Ser Ser Ala Ala Thr Pro Asn Leu Gly Pro Val  
 435 440 445  
 Gly Ser Gly Met Leu Asn Asn His Gly His Ala Val Pro Ala Asn Ser  
 450 455 460  
 Glu Met Thr Ser Ser His Gly Thr Gln Ser Met Val Ser Gly Ser His  
 465 470 475 480  
 Cys Thr Pro Pro Pro Tyr His Ala Asp Pro Ser Leu Val Arg Thr  
 485 490 495  
 Trp Gly Pro

(2) INFORMATION FOR SEQ ID NO: 5:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2156 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(11) MOLECULE TYPE: cDNA

(12) ORIGINAL SOURCE:

(A) ORGANISM: Homo sapiens

(13) FEATURE:

(A) NAME/KEY: CDS

(B) LOCATION: 33..1940

(14) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

GCGAGCTGCC	CTCGGAGGCC	GGCGTGGGGA	AG	ATG	GCC	CAG	TCC	ACC	GCC	ACC	53					
				Met	Ala	Gln	Ser	Thr	Ala	Thr						
				1				5								
TCC	CCT	GAT	GGG	GGC	ACC	ACG	TTT	GAG	CAC	CTC	TGG	AGC	TCT	CTG	GAA	101
Ser	Pro	Asp	Gly	Gly	Thr	Thr	Phe	Glu	His	Leu	Trp	Ser	Ser	Leu	Glu	
			10				15					20				
CCA	GAC	AGC	ACC	TAC	TTC	GAC	CTT	CCC	CAG	TCA	AGC	CGG	GGG	AAT	AAT	149
Pro	Asp	Ser	Thr	Tyr	Phe	Asp	Leu	Pro	Gln	Ser	Ser	Arg	Gly	Asn	Asn	
			25			30					35					
GAG	GTG	GTG	GGC	GGA	ACG	GAT	TCC	AGC	ATG	GAC	GTC	TTC	CAC	CTG	GAG	197
Glu	Val	Val	Gly	Gly	Thr	Asp	Ser	Ser	Met	Asp	Val	Phe	His	Leu	Glu	
	40				45					50				55		
GGC	ATG	ACT	ACA	TCT	GTC	ATG	GCC	CAG	TTC	AAT	CTG	CTG	AGC	AGC	ACC	245
Gly	Met	Thr	Thr	Ser	Val	Met	Ala	Gln	Phe	Asn	Leu	Leu	Ser	Ser	Thr	
			60					65					70			
ATG	GAC	CAG	ATG	AGC	AGC	CGC	GCG	GCC	TCG	GCC	AGC	CCC	TAC	ACC	CCA	293
Met	Asp	Gln	Met	Ser	Ser	Arg	Ala	Ala	Ser	Ala	Ser	Pro	Tyr	Thr	Pro	
			75					80					85			
GAG	CAC	GCC	GCC	AGC	GTG	CCC	ACC	CAC	TCG	CCC	TAC	GCA	CAA	CCC	AGC	341
Glu	His	Ala	Ala	Ser	Val	Pro	Thr	His	Ser	Pro	Tyr	Ala	Gln	Pro	Ser	
		90				95					100					
TCC	ACC	TTC	GAC	ACC	ATG	TCG	CGG	GCG	CCT	GTC	ATC	CCC	TCC	AAC	ACC	389
Ser	Thr	Phe	Asp	Thr	Met	Ser	Pro	Ala	Pro	Val	Ile	Pro	Ser	Asn	Thr	
	105				110					115						
GAC	TAC	CCC	GGA	CCC	CAC	CAC	TTT	GAG	GTC	ACT	TTC	CAG	CAG	TCC	AGC	437
Asp	Tyr	Pro	Gly	Pro	His	His	Phe	Glu	Val	Thr	Phe	Gln	Gln	Ser	Ser	
	120				125				130					135		
ACG	GCC	AAG	TCA	GCC	ACC	TGG	ACG	TAC	TCC	CGG	CTC	TTG	AAG	AAA	CTC	485
Thr	Ala	Lys	Ser	Ala	Thr	Trp	Thr	Tyr	Ser	Pro	Leu	Leu	Lys	Lys	Leu	
			140						145				150			
TAC	TGC	CAG	ATC	GCC	AAG	ACA	TGC	CCC	ATC	CAG	ATC	AAG	GTG	TCC	ACC	533
Tyr	Cys	Gln	Ile	Ala	Lys	Thr	Cys	Pro	Ile	Gln	Ile	Lys	Val	Ser	Thr	
		155					160					165				
CCG	CCA	CCC	CCA	GGC	ACT	GCC	ATC	CGG	GCC	ATG	CCT	GTT	TAC	AAG	AAA	581
Pro	Pro	Pro	Pro	Gly	Thr	Ala	Ile	Arg	Ala	Met	Pro	Val	Tyr	Lys	Lys	
		170				175					180					
GCG	GAG	CAC	GTG	ACC	GAC	GTG	GTG	AAA	CGC	TGC	CCC	AAC	CAC	GAG	CTC	629
Ala	Glu	His	Val	Thr	Asp	Val	Val	Lys	Arg	Cys	Pro	Asn	His	Glu	Leu	
	185				190					195						
GGG	AGG	GAC	TTC	AAC	GAA	GGA	CAG	TCT	GCT	CCA	GCC	AGC	CAC	CTC	ATC	677
Gly	Arg	Asp	Phe	Asn	Glu	Gly	Gln	Ser	Ala	Pro	Ala	Ser	His	Leu	Ile	
	200				205					210				215		

CGC GTG GAA GGC AAT AAT CTC TCG CAG TAT GTG GAT GAC CCT GTC ACC Arg Val Glu Gly Asn Asn Leu Ser Gln Tyr Gln Val Asp Asp Pro Val Thr 220 225 230	725
GGC AGG CAG AGC GTC GTG GTG CCC TAT GAG CCA CCA CAG GTG GGG ACG Gly Arg Gln Ser Val Val Val Pro Tyr Gln Pro Gln Val Gly Thr 235 240 245	773
GAA TTC ACC ACC ATC CTC TAC AAC TTC ATG TGT AAC AGC AGC TGT GTA Glu Phe Thr Thr Ile Leu Tyr Asn Phe Met Cys Asn Ser Ser Cys Val 250 255 260	821
GGG GGC ATG AAC CGG CGG CCC ATC CTC ATC ATC ACC CTG GAG ATG Gly Gly Met Asn Arg Arg Pro Ile Leu Ile Ile Ile Thr Leu Glu Met 265 270	869
CGG GAT GGG CAG GTG CTG GGC CGC CGG TCC TTT GAG GGC CGC ATC TGC Arg Asp Gly Gln Val Leu Gly Arg Arg Ser Phe Glu Gly Arg Ile Cys 280 285 290 295	917
GCC TGT CCT GGC CGC GAC CGA AAA GCT GAT GAG GAC CAC TAC CGG GAG Ala Cys Pro Gly Arg Asp Arg Lys Ala Asp Glu Asp His Tyr Arg Glu 300 305 310	965
CAG CAG GCC CTG AAC GAG AGC TCC GCC AAG AAC GGG GCC GCC AGC AAG Gln Gln Ala Leu Asn Glu Ser Ser Ala Lys Asn Gly Ala Ala Ser Lys 315 320 325	1013
CGT GCC TTC AAG CAG AGC CCC CCT GCC GTC CCC GCC CTT GGT GCC GGT Arg Ala Phe Lys Gln Ser Pro Pro Ala Val Pro Ala Leu Gly Ala Gly 330 335 340	1061
GTG AAG AAG CGG CGG CAT GGA GAC GAG GAC ACG TAC TAC CTT CAG GTG Val Lys Lys Arg Arg His Gly Asp Glu Asp Thr Tyr Leu Gln Val 345 350 355	1109
CGA GGC CGG GAG AAC TTT GAG ATC CTG ATG AAG CTG AAA GAG AGC CTG Arg Gly Arg Glu Asn Phe Glu Ile Leu Met Lys Leu Lys Glu Ser Leu 360 365 370 375	1157
GAG CTG ATG GAG TTG GTG CCG CAG CCA CTG GTG GAC TCC TAT CGG CAG Glu Leu Met Glu Leu Val Pro Gln Pro Leu Val Asp Ser Tyr Arg Gln 380 385 390	1205
CAG CAG CAG CTC CTA CAG AGG CCG AGT CAC CTA CAG CCC CCG TCC TAC Gln Gln Gln Leu Leu Gln Arg Pro Pro Ser His Leu Gln Pro Pro Ser Tyr 395 400 405	1253
GGG CCG GTC CTC TCG CCC ATG AAC AAG GTG CAC GGG GGC ATG AAC AAG Gly Pro Val Leu Ser Pro Met Asn Lys Val His Gly Gly Met Asn Lys 410 415 420	1301
CTG CCC TCC GTC AAC CAG CTG GTG GGC CAG CCT CCC CCG CAC AGT TCG Leu Pro Ser Val Asn Gln Leu Val Gly Gln Pro Pro Pro His Ser Ser 425 430 435	1349
GCA GGT ACA CCC AAC CTG GGG CCC GTG GGC CCC GGG ATG CTC AAC AAC Ala Ala Thr Pro Asn Leu Gly Pro Val Gly Pro Gly Met Leu Asn Asn 440 445 450 455	1397
CAT GGC CAC GCA GTG CCA GCC AAC GGC GAG ATG AGC AGC AGC CAC AGC His Gly His Ala Val Pro Ala Asn Gly Glu Met Ser Ser Ser Ser Ser 460 465 470	1445
GCC CAG TCC ATG GTC TCG GGG TCC CAC TGC ACT CCG CCA CCC CCC TAC Ala Gln Ser Met Val Ser Gly Ser His Cys Thr Pro Pro Pro Pro Tyr 475 480 485	1493
CAC GCC GAC CCC AGC CTC GTC AGT TTT TTA ACA GGA TTG GGG TGT CCA His Ala Asp Pro Ser Leu Val Ser Phe Leu Thr Gly Leu Gly Cys Pro 490 495 500	1541

AAC TGC ATC GAG TAT TTC ACC TCC GAA GGG TTA CAG AGC ATT TAC CAC 1589  
Asn Cys Ile Glu Tyr Phe Thr Ser Gln Gly Leu Gln Ser Ile Tyr His  
505 510 515  
CTG CAG AAC CTG ACC ATT GAG GAC CTG GGG GCC CTG AAG ATC CCC GAG 1637  
Leu Gln Asn Leu Thr Ile Glu Asp Leu Gly Ala Leu Lys Ile Pro Glu  
520 525 530 535  
CAG TAC CGC ATG ACC ATC TGG CGG GGC CTG CAG GAC CTG AAG CAG GGC 1685  
Gln Tyr Arg Met Thr Ile Trp Arg Gly Leu Gln Asp Leu Lys Gln Gly  
540 545 550  
CAC GAC TAC AGC ACC GCG CAG CAG CTG CTC GCG TCT AGC AAC GCG GCC 1733  
His Asp Tyr Ser Thr Ala Gln Gln Leu Leu Arg Ser Ser Asn Ala Ala  
555 560 565  
ACC ATC TCC ATC GGC GGC TCA GGG GAA CTG CAG CGC CAG CGG GTC ATG 1791  
Thr Ile Ser Ile Gly Gly Ser Gly Glu Leu Gln Arg Gln Arg Val Met  
570 575 580  
GAG GCC GTG CAC TTC CGC GTG CGC CAC ACC ATC ACC ATC CCC AAC CGC 1829  
Glu Ala Val His Phe Arg Val Arg His Thr Ile Thr Ile Pro Asn Arg  
585 590 595  
GGC GGC CCA GGC GGC GGC CCT GAC GAG TGG GCG GAC TTC GGC TTC GAC 1877  
Gly Gly Pro Gly Gly Gly Pro Asp Glu Trp Ala Asp Phe Gly Phe Asp  
600 605 610 615  
CTG CCC GAC TGC AAG GCC CGC AAG CAG CCC ATC AAG GAG GAG TTC ACG 1925  
Leu Pro Asp Cys Lys Ala Arg Lys Gln Pro Ile Lys Glu Glu Phe Thr  
620 625 630  
GAG GCC GAG ATC CAC TGAGGGGCTC GCCTGGCTGC AGCGTGGGCC ACCGGCCAGA 1980  
Glu Ala Glu Ile His  
635  
GACCCAAAGCT GCCTCCCTC TCCTTCCTGT GTGTCCAAAA CTGCCTCAGG AGGCAGGACC 2040  
TTGGGGCTGT GCCCGGGGAA AGGCAAGGTC CGGCCCATCC CCAGGCACCT CACAGGCCCC 2100  
AGGAAGGCC CAGCCACCGA AGCCGCTGT GGACAGCCTG AGTCACCTGC AGAACC 2156

## (2) INFORMATION FOR SEQ ID NO: 6:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 636 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

## (11) MOLECULE TYPE: protein

## (X1) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

Met Ala Gln Ser Thr Ala Thr Ser Pro Asp Gly Gly Thr Thr Phe Glu  
1 5 10 15  
His Leu Trp Ser Ser Leu Glu Pro Asp Ser Thr Tyr Phe Asp Leu Pro  
20 25 30  
Gln Ser Ser Arg Gly Asn Asn Glu Val Val Gly Gly Thr Asp Ser Ser  
35 40 45  
Met Asp Val Phe His Leu Glu Gly Met Thr Thr Ser Val Met Ala Gln  
50 55 60  
Phe Asn Leu Leu Ser Ser Thr Met Asp Gln Met Ser Ser Arg Ala Ala  
65 70 75 80  
Ser Ala Ser Pro Tyr Thr Pro Glu His Ala Ala Ser Val Pro Thr His  
85 90 95

Ser Pro Tyr Ala Gln Pro Ser Ser Thr Phe Asp Thr Met Ser Pro Ala  
 100 105 110  
 Pro Val Ile Pro Ser Asn Thr Asp Tyr Pro Gly Pro His His Phe Glu  
 115 120 125  
 Val Thr Phe Gln Gln Ser Ser Thr Ala Lys Ser Ala Thr Trp Thr Tyr  
 130 135 140  
 Ser Pro Leu Leu Lys Lys Leu Tyr Cys Gln Ile Ala Lys Thr Cys Pro  
 145 150 155 160  
 Ile Gln Ile Lys Val Ser Thr Pro Pro Pro Pro Gly Thr Ala Ile Arg  
 165 170 175  
 Ala Met Pro Val Tyr Lys Lys Ala Glu His Val Thr Asp Val Val Lys  
 180 185 190  
 Arg Cys Pro Asn His Glu Leu Gly Arg Asp Phe Asn Glu Gly Gln Ser  
 195 200 205  
 Ala Pro Ala Ser His Leu Ile Arg Val Glu Gly Asn Asn Leu Ser Gln  
 210 215 220  
 Tyr Val Asp Asp Pro Val Thr Gly Arg Gln Ser Val Val Val Pro Tyr  
 225 230 235 240  
 Glu Pro Pro Gln Val Gly Thr Glu Phe Thr Thr Ile Leu Tyr Asn Phe  
 245 250 255  
 Met Cys Asn Ser Ser Cys Val Gly Gly Met Asn Arg Arg Pro Ile Leu  
 260 265 270  
 Ile Ile Ile Thr Leu Glu Met Arg Asp Gly Gln Val Leu Gly Arg Arg  
 275 280 285  
 Ser Phe Glu Gly Arg Ile Cys Ala Cys Pro Gly Arg Asp Arg Lys Ala  
 290 295 300  
 Asp Glu Asp His Tyr Arg Glu Gln Gln Ala Leu Asn Glu Ser Ser Ala  
 305 310 315 320  
 Lys Asn Gly Ala Ala Ser Lys Arg Ala Phe Lys Gln Ser Pro Pro Ala  
 325 330 335  
 Val Pro Ala Leu Gly Ala Gly Val Lys Lys Arg Arg His Gly Asp Glu  
 340 345 350  
 Asp Thr Tyr Tyr Leu Gln Val Arg Gly Arg Glu Asn Phe Glu Ile Leu  
 355 360 365  
 Met Lys Leu Lys Glu Ser Leu Glu Leu Met Glu Leu Val Pro Gln Pro  
 370 375 380  
 Leu Val Asp Ser Tyr Arg Gln Gln Gln Gln Leu Leu Gln Arg Pro Ser  
 385 390 395 400  
 His Leu Gln Pro Pro Ser Tyr Gly Pro Val Leu Ser Pro Met Asn Lys  
 405 410 415  
 Val His Gly Gly Met Asn Lys Leu Pro Ser Val Asn Gln Leu Val Gly  
 420 425 430  
 Gln Pro Pro Pro His Ser Ser Ala Ala Thr Pro Asn Leu Gly Pro Val  
 435 440 445  
 Gly Pro Gly Met Leu Asn Asn His Gly His Ala Val Pro Ala Asn Gly  
 450 455 460  
 Glu Met Ser Ser Ser His Ser Ala Gln Ser Met Val Ser Gly Ser His  
 465 470 475 480

Cys Thr Pro Pro Pro Tyr His Ala Asp Pro Ser Leu Val Ser Phe  
 485 490 495  
 Leu Thr Gly Leu Gly Cys Pro Asn Cys Ile Glu Tyr Phe Thr Ser Gln  
 500 505 510  
 Gly Leu Gln Ser Ile Tyr His Leu Gln Asn Leu Thr Ile Glu Asp Leu  
 515 520 525  
 Gly Ala Leu Lys Ile Pro Glu Gln Tyr Arg Met Thr Ile Trp Arg Gly  
 530 535 540  
 Leu Gln Asp Leu Lys Gln Gly His Asp Tyr Ser Thr Ala Gln Gln Leu  
 545 550 555 560  
 Leu Arg Ser Ser Asn Ala Ala Thr Ile Ser Ile Gly Gly Ser Gly Glu  
 565 570 575  
 Leu Gln Arg Gln Arg Val Met Glu Ala Val His Phe Arg Val Arg His  
 580 585 590  
 Thr Ile Thr Ile Pro Asn Arg Gly Gly Pro Gly Gly Gly Pro Asp Glu  
 595 600 605  
 Trp Ala Asp Phe Gly Phe Asp Leu Pro Asp Cys Lys Ala Arg Lys Gln  
 610 615 620  
 Pro Ile Lys Glu Glu Phe Thr Glu Ala Glu Ile His  
 625 630 635

## (2) INFORMATION FOR SEQ ID NO: 7:

- (1) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 2040 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(11) MOLECULE TYPE: cDNA

(v1) ORIGINAL SOURCE:

(A) ORGANISM: Mus musculus

(1x) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 124..1890

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

TGATCTCCCT GTGGCTGCA GGGGACTGAG CCAGGGAGTA GATGCCCTGA GACCCCAAGG	60
GACACCCAAG GAAACCTTGC TGGCTTIGAG AAAGGGATCG TCTCTCTCT GCCCAAGAGA	120
AGC ATG TGT ATG GGC CCT GTG TAT GAA TCC TTG GGG CAG GCC CAG TTC	188
Met Cys Met Gly Pro Val Tyr Glu Ser Leu Gly Gln Ala Gln Phe	
1 5 10 15	
AAT TTG CTC AGC AGT GCC ATG GAC CAG ATG GGC AGC CGT GCG GCC CCG	216
Asn Leu Leu Ser Ser Ala Met Asp Gln Met Gly Ser Arg Ala Ala Pro	
20 25 30	
GGG AGC CCC TAC ACC CCG GAG CAC GCC GCC AGC GCG CCC ACC CAC TCG	264
Ala Ser Pro Tyr Thr Pro Glu His Ala Ala Ser Ala Pro Thr His Ser	
35 40 45	
CCC TAC GCG CAG CCC AGC TCC ACC TTC GAC ACC ATG TCT CCG GCG CCT	312
Pro Tyr Ala Gln Pro Ser Ser Thr Phe Asp Thr Met Ser Pro Ala Pro	
50 55 60	
GTC ATC CCT TCC AAT ACC GAC TAC CCC GGC CCC CAC CAC TTC GAG GTC	360

Val	Ile	Pro	Ser	Asn	Thr	Asp	Tyr	Pro	Gly	Pro	His	His	Phe	Glu	Val	
65						70					75					
ACC	TTC	CAG	CAG	TCG	AGC	ACT	GCC	AAG	TCG	GCC	ACC	TGG	ACA	TAC	TCC	408
Thr	Phe	Gln	Gln	Ser	Thr	Ala	Lys	Ser	Ala	Thr	Trp	Thr	Tyr	Ser		
80				85					90					95		
CCA	CTC	TTC	AAG	AAG	TTG	TAC	TGT	CAG	ATT	GCT	AAG	ACA	TGC	CCC	ATC	456
Pro	Leu	Leu	Lys	Lys	Leu	Tyr	Cys	Gln	Ile	Ala	Lys	Thr	Cys	Pro	Ile	
			100					105					110			
CAG	ATC	AAA	GTG	TCC	ACA	CCA	CCA	CCC	CCG	GGC	ACG	GCC	ATC	CGG	GCC	504
Gln	Ile	Lys	Val	Ser	Thr	Pro	Pro	Pro	Pro	Gly	Thr	Ala	Ile	Arg	Ala	
			115					120				125				
ATG	CCT	GTG	TAC	AAG	AAG	GCA	GAG	CAT	GTG	ACC	GAC	ATT	GTT	AAG	CGC	552
Met	Pro	Val	Tyr	Lys	Lys	Ala	Glu	His	Val	Thr	Asp	Ile	Val	Lys	Arg	
		130					135					140				
TGC	CCC	AAC	CAC	GAG	CTT	GGA	AGG	GAC	TTC	AAT	GAA	GGA	CAG	TCT	GCC	600
Cys	Pro	Asn	His	Glu	Leu	Gly	Arg	Asp	Phe	Asn	Glu	Gly	Gln	Ser	Ala	
		145				150					155					
CCG	GCT	AGC	CAC	CTC	ATC	CGT	GTA	GAA	GGC	AAC	AAC	CTC	GCC	CAG	TAC	648
Pro	Ala	Ser	His	Leu	Ile	Arg	Val	Glu	Gly	Asn	Asn	Leu	Ala	Gln	Tyr	
		160			165				170					175		
GTG	GAT	GAC	CCT	GTC	ACC	GGA	AGG	CAG	AGT	GTG	GTT	GTG	CCG	TAT	GAA	696
Val	Asp	Asp	Pro	Val	Thr	Gly	Arg	Gln	Ser	Val	Val	Val	Pro	Tyr	Glu	
			180					185					190			
CCC	CCA	CAG	GTG	GGA	ACA	GAA	TTT	ACC	ACC	ATC	CTG	TAC	AAC	TTC	ATG	744
Pro	Pro	Gln	Val	Gly	Thr	Glu	Phe	Thr	Thr	Ile	Leu	Tyr	Asn	Phe	Met	
			195					200				205				
TGT	AAC	AGC	AGC	TGT	GTG	GGG	GGC	ATG	AAT	CGG	AGG	CCC	ATC	CTT	GTC	792
Cys	Asn	Ser	Ser	Cys	Val	Gly	Gly	Met	Asn	Arg	Arg	Pro	Ile	Leu	Val	
		210				215						220				
ATC	ATC	ACC	CTG	GAG	ACC	CGG	GAT	GGA	CAG	GTC	CTG	GSC	CGC	CGG	TCT	840
Ile	Ile	Thr	Leu	Glu	Thr	Arg	Asp	Gly	Gln	Val	Leu	Gly	Arg	Arg	Ser	
			225			230					235					
TTC	GAG	GGT	CGC	ATC	TGT	GCC	TGT	CCT	GGC	CGT	GAC	CGC	AAA	CCT	GAT	888
Phe	Glu	Gly	Arg	Ile	Cys	Ala	Cys	Pro	Gly	Arg	Asp	Arg	Lys	Ala	Asp	
		240			245				250				255			
GAA	GAC	CAT	TAC	CGG	GAG	CAA	CAG	GCT	CTG	AAT	GAA	AGT	ACC	ACC	AAA	936
Glu	Asp	His	Tyr	Arg	Glu	Gln	Gln	Ala	Leu	Asn	Glu	Ser	Thr	Thr	Lys	
			260					265					270			
AAT	GGA	GCT	GCC	AGC	AAA	CGT	GCA	TTC	AAG	CAG	AGC	CCC	CCT	GCC	ATC	984
Asn	Gly	Ala	Ala	Ser	Lys	Arg	Ala	Phe	Lys	Gln	Ser	Pro	Pro	Ala	Ile	
			275				280						285			
CCT	GCC	CTG	GGT	ACC	AAC	GTG	AAG	AAG	AGA	CGC	CAC	GGG	GAC	GAG	GAC	1032
Pro	Ala	Leu	Gly	Thr	Asn	Val	Lys	Lys	Arg	Arg	His	Gly	Asp	Glu	Asp	
		290				295						300				
ATG	TTC	TAC	ATG	CAC	GTG	CGA	GGC	CGG	GAG	AAC	TTT	GAG	ATC	TTG	ATG	1080
Met	Phe	Tyr	Met	His	Val	Arg	Gly	Arg	Glu	Asn	Phe	Glu	Ile	Leu	Met	
		305				310					315					
AAA	GTC	AAG	GAG	AGC	CTA	GAA	CTG	ATG	GAG	CTT	GTG	CCC	CAG	CCT	TTG	1128
Lys	Val	Lys	Glu	Ser	Leu	Glu	Leu	Met	Glu	Leu	Val	Pro	Gln	Pro	Leu	
		320			325				330					335		
GTT	GAC	TCC	TAT	CGA	CAG	CAG	CAG	CAG	CAG	CAG	CTC	CTA	CAG	AGG	CCG	1176
Val	Arg	Ser	Tyr	Arg	Gln	Gln	Gln	Gln	Gln	Leu	Leu	Gln	Arg	Pro		
			340				345						350			
AGT	CAC	CTG	CAG	CCT	CCA	TCC	TAT	GGG	CCC	GTG	CTC	TCC	CCA	ATG	AAC	1224

Ser His Leu Gln Pro Pro Ser Tyr Gly Pro Val Leu Ser Pro Met Asn	
355	360
365	
AAG GTA CAC GGT GGT GTC AAC AAA CTG CCC TCC GTC AAC CAG CTG GTG	1272
Lys Val His Gly Gly Val Asn Lys Leu Pro Ser Val Asn Gln Leu Val	
370	375
380	
GGC CAG CCT CCC CCG CAC AGC TCA GCA GCT GGG CCC AAC CTG GGG CCC	1320
Gly Gln Pro Pro Pro His Ser Ser Ala Ala Gly Pro Asn Leu Gly Pro	
385	390
395	
ATG GGC TCC GGG ATG CTC AAC AGC CAC GGC CAC AGC ATG CCG GCC AAT	1368
Met Gly Ser Gly Met Leu Asn Ser His Gly His Ser Met Pro Ala Asn	
400	405
410	415
GGT GAG ATG AAT GGA GGC CAC AGC TCC CAG ACC ATG GTT TCG GGA TCC	1416
Gly Glu Met Asn Gly Gly His Ser Ser CAG Thr Met Val Ser Gly Ser	
420	425
430	
CAC TGC ACC CCG CCA CCC CCC TAT CAT GCA GAC CCC AGC CTC GTC AGT	1464
His Cys Thr Pro Pro Pro Tyr His Ala Asp Pro Ser Leu Val Ser	
435	440
445	
TTT TTG ACA GGG TTG GGG TGT CCA AAC TGC ATC GAG TGC TTC ACT TCC	1512
Phe Leu Thr Gly Leu Gly Cys Pro Asn Cys Ile Glu Cys Phe Thr Ser	
450	455
460	
CAA GGG TTG CAG AGC ATC TAC CAC CTG CAG AAC CTT ACC ATC GAG GAC	1560
Gln Gly Leu Gln Ser Ile Tyr His Leu Gln Asn Leu Thr Ile Glu Asp	
465	470
475	
CTT GGG GCT CTG AAG GTC CCT GAC CAG TAC CGT ATG ACC ATC TGG AGG	1608
Leu Gly Ala Leu Lys Val Pro Asp Gln Tyr Arg Met Thr Ile Trp Arg	
480	485
490	495
GGC CTA CAG GAC CTG AAG CAG AGC CAT GAC TGC GGC CAG CAA CTG CTA	1656
Gly Leu Gln Asp Leu Lys Gln Ser His Asp Cys Gly Gln Gln Leu Leu	
500	505
510	
CGC TCC AGC AGC AAC GCG GCC ACC ATC TCC ATC GGC GGC TCT GGC GAG	1704
Arg Ser Ser Ser Asn Ala Ala Thr Ile Ser Ile Gly Gly Ser Gly Glu	
515	520
525	
CTG CAG CGG CAG CGG GTC ATG GAA GCC GTG CAT TTC CGT GTG CGC CAC	1752
Leu Gln Arg Gln Arg Val Met Glu Ala Val His Phe Arg Val Arg His	
530	535
540	
ACC ATC ACA ATC CCC AAC CGT GGA GGC GCA GGT GCG GTG ACA GGT CCC	1800
Thr Ile Thr Ile Pro Asn Arg Gly Gly Ala Gly Ala Val Thr Gly Pro	
545	550
555	
GAC GAG TGG GCG GAC TTT GGC TTT GAC CTG CCT GAC TGC AAG TCC CGT	1848
Asp Glu Trp Ala Asp Phe Gly Phe Asp Leu Pro Asp Cys Lys Ser Arg	
560	565
570	575
AAG CAG CCC ATC AAA GAG GAG TTC ACA GAG ACA GAG AGC CAC	1890
Lys Gln Pro Ile Lys Glu Glu Phe Thr Glu Thr Glu Ser His	
580	585
590	
TGAGGAACGT ACCCTCTTCT CCGTCCTTC CTCTGTGAGA AACTGCTCTT GGAAGTGGGA	1950
CGTGTGGCT GTGCCACAG AAACCAGCAA GGACCTTCTG CCGGATGCCA TTCCTGAAGG	2010
GAAGTCGCTC ATGAACATAAC TCCTCTTGG	2040

(2) INFORMATION FOR SEQ ID NO: 8:

- (1) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 589 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(11) MOLECULE TYPE: protein

(12) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

```

Met Cys Met Gly Pro Val Tyr Glu Ser Leu Gly Gln Ala Gln Phe Asn
 1           5           10           15
Leu Leu Ser Ser Ala Met Asp Gln Met Gly Ser Arg Ala Ala Pro Ala
          20           25           30
Ser Pro Tyr Thr Pro Glu His Ala Ser Ala Pro Thr His Ser Pro
          35           40           45
Tyr Ala Gln Pro Ser Ser Thr Phe Asp Thr Met Ser Pro Ala Pro Val
          50           55           60
Ile Pro Ser Asn Thr Asp Tyr Pro Gly Pro His His Phe Glu Val Thr
 65           70           75           80
Phe Gln Gln Ser Ser Thr Ala Lys Ser Ala Thr Trp Thr Tyr Ser Pro
          85           90           95
Leu Leu Lys Lys Leu Tyr Cys Gln Ile Ala Lys Thr Cys Pro Ile Gln
          100          105          110
Ile Lys Val Ser Thr Pro Pro Pro Gly Thr Ala Ile Arg Ala Met
          115          120          125
Pro Val Tyr Lys Lys Ala Glu His Val Thr Asp Ile Val Lys Arg Cys
          130          135          140
Pro Asn His Glu Leu Gly Arg Asp Phe Asn Glu Gly Gln Ser Ala Pro
          145          150          155          160
Ala Ser His Leu Ile Arg Val Glu Gly Asn Asn Leu Ala Gln Tyr Val
          165          170          175
Asp Asp Pro Val Thr Gly Arg Gln Ser Val Val Val Pro Tyr Glu Pro
          180          185          190
Pro Gln Val Gly Thr Glu Phe Thr Thr Ile Leu Tyr Asn Phe Met Cys
          195          200          205
Asn Ser Ser Cys Val Gly Gly Met Asn Arg Arg Pro Ile Leu Val Ile
          210          215          220
Ile Thr Leu Glu Thr Arg Asp Gly Gln Val Leu Gly Arg Arg Ser Phe
          225          230          235          240
Glu Gly Arg Ile Cys Ala Cys Pro Gly Arg Asp Arg Lys Ala Asp Glu
          245          250          255
Asp His Tyr Arg Glu Gln Gln Ala Leu Asn Glu Ser Thr Thr Lys Asn
          260          265          270
Gly Ala Ala Ser Lys Arg Ala Phe Lys Gln Ser Pro Pro Ala Ile Pro
          275          280          285
Ala Leu Gly Thr Asn Val Lys Lys Arg Arg His Gly Asp Glu Asp Met
          290          295          300
Phe Tyr Met His Val Arg Gly Arg Glu Asn Phe Glu Ile Leu Met Lys
          305          310          315          320
Val Lys Glu Ser Leu Glu Leu Met Glu Leu Val Pro Gln Pro Leu Val
          325          330          335
Asp Ser Tyr Arg Gln Gln Gln Gln Gln Leu Leu Gln Arg Pro Ser
          340          345          350
His Leu Gln Pro Pro Ser Tyr Gly Pro Val Leu Ser Pro Met Asn Lys

```

(11) MOLECULE TYPE: protein

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

```

Met Cys Met Gly Pro Val Tyr Glu Ser Leu Gly Gln Ala Gln Phe Asn
 1           5           10
Leu Leu Ser Ser Ala Met Asp Gln Met Gly Ser Arg Ala Ala Pro Ala
      20           25           30
Ser Pro Tyr Thr Pro Glu His Ala Ala Ser Ala Pro Thr His Ser Pro
      35           40           45
Tyr Ala Gln Pro Ser Ser Thr Phe Asp Thr Met Ser Pro Ala Pro Val
      50           55           60
Ile Pro Ser Asn Thr Asp Tyr Pro Gly Pro His His Phe Glu Val Thr
      65           70           75           80
Phe Gln Gln Ser Ser Thr Ala Lys Ser Ala Thr Trp Thr Tyr Ser Pro
      85           90           95
Leu Leu Lys Lys Leu Tyr Cys Gln Ile Ala Lys Thr Cys Pro Ile Gln
      100           105           110
Ile Lys Val Ser Thr Pro Pro Pro Gly Thr Ala Ile Arg Ala Met
      115           120           125
Pro Val Tyr Lys Lys Ala Glu His Val Thr Asp Ile Val Lys Arg Cys
      130           135           140
Pro Asn His Glu Leu Gly Arg Asp Phe Asn Glu Gly Gln Ser Ala Pro
      145           150           155           160
Ala Ser His Leu Ile Arg Val Glu Gly Asn Leu Ala Gln Tyr Val
      165           170           175
Asp Asp Pro Val Thr Gly Arg Gln Ser Val Val Val Pro Tyr Glu Pro
      180           185           190
Pro Gln Val Gly Thr Glu Phe Thr Thr Ile Leu Tyr Asn Phe Met Cys
      195           200           205
Asn Ser Ser Cys Val Gly Gly Met Asn Arg Arg Pro Ile Leu Val Ile
      210           215           220
Ile Thr Leu Glu Thr Arg Asp Gly Gln Val Leu Gly Arg Arg Ser Phe
      225           230           235           240
Glu Gly Arg Ile Cys Ala Cys Pro Gly Arg Asp Arg Lys Ala Asp Glu
      245           250           255
Asp His Tyr Arg Glu Gln Gln Ala Leu Asn Glu Ser Thr Thr Lys Asn
      260           265           270
Gly Ala Ala Ser Lys Arg Ala Phe Lys Gln Ser Pro Pro Ala Ile Pro
      275           280           285
Ala Leu Gly Thr Asn Val Lys Lys Arg Arg His Gly Asp Glu Asp Met
      290           295           300
Phe Tyr Met His Val Arg Gly Arg Glu Asn Phe Glu Ile Leu Met Lys
      305           310           315           320
Val Lys Glu Ser Leu Glu Leu Met Glu Leu Val Pro Gln Pro Leu Val
      325           330           335
Asp Ser Tyr Arg Gln Gln Gln Gln Gln Leu Leu Gln Arg Pro Ser
      340           345           350
His Leu Gln Pro Pro Ser Tyr Gly Pro Val Leu Ser Pro Met Asn Lys

```

355	360	365
Val His Gly Gly Val Asn Lys Leu Pro Ser Val Asn Gln Leu Val Gly		
370	375	380
Gln Pro Pro Pro His Ser Ser Ala Ala Gly Pro Asn Leu Gly Pro Met		
385	390	395
Gly Ser Gly Met Leu Asn Ser His Gly His Ser Met Pro Ala Asn Gly		
405	410	415
Glu Met Asn Gly Gly His Ser Ser Gln Thr Met Val Ser Gly Ser His		
420	425	430
Cys Thr Pro Pro Pro Tyr His Ala Asp Pro Ser Leu Val Ser Phe		
435	440	445
Leu Thr Gly Leu Gly Cys Pro Asn Cys Ile Glu Cys Phe Thr Ser Gln		
450	455	460
Gly Leu Gln Ser Ile Tyr His Leu Gln Asn Leu Thr Ile Glu Asp Leu		
465	470	475
Gly Ala Leu Lys Val Pro Asp Gln Tyr Arg Met Thr Ile Trp Arg Gly		
485	490	495
Leu Gln Asp Leu Lys Gln Ser His Asp Cys Gly Gln Gln Leu Leu Arg		
500	505	510
Ser Ser Ser Asn Ala Ala Thr Ile Ser Ile Gly Gly Ser Gly Glu Leu		
515	520	525
Gln Arg Gln Arg Val Met Glu Ala Val His Phe Arg Val Arg His Thr		
530	535	540
Ile Thr Ile Pro Asn Arg Gly Gly Ala Gly Ala Val Thr Gly Pro Asp		
545	550	555
Glu Trp Ala Asp Phe Gly Phe Asp Leu Pro Asp Cys Lys Ser Arg Lys		
565	570	575
Gln Pro Ile Lys Glu Glu Phe Thr Glu Thr Glu Ser His		
580	585	

## (2) INFORMATION FOR SEQ ID NO: 9:

- (1) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 758 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear
- (11) MOLECULE TYPE: cDNA
- (VI) ORIGINAL SOURCE:
  - (A) ORGANISM: Mus musculus
- (1X) FEATURE:
  - (A) NAME/KEY: CDS
  - (B) LOCATION: 389..757

## (X1) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

TGGTCCCGCT TCGACCAAGA CTCGGGCTAC CAGCTTGGCG GCCCGCGGGA GGAGGAGACC	60
CCGCTGGGGC TAGCTGGGCG ACGCGCGCCA AGCGCGGGCG GGAAGGAGGC GGGAGGAGCG	120
GGGCCCCGAGA CCGCGACTCG GGCAGAGCCA GCTGGGGAGG CGGGGCGCGC GTGGGAGCCA	180
GGGGCCCGGG TGGCCGGCCC TCCTCGGCCA CGGCTGAGTG CCGCGGTGCG CTTCGCCCGC	240

(2) INFORMATION FOR SEQ ID NO: 10:

(11) MOLECULE TYPE: protein

Met Ser Gly Ser Val Gly Glu Met Ala Gln Thr Ser Ser Ser Ser Ser  
1 5 10 15  
Ser Thr Phe Glu His Leu Trp Ser Ser Leu Glu Pro Asp Ser Thr Tyr  
20 25 30  
Phe Asp Leu Pro Gln Pro Ser Gln Gly Thr Ser Glu Ala Ser Gly Ser  
35 40 45  
Glu Glu Ser Asn Met Asp Val Phe His Leu Gln Gly Met Ala Gln Phe  
50 55 60  
Asn Leu Ser Ser Ala Met Asp Gln Met Gly Ser Arg Ala Ala Pro  
65 70 75 80  
Ala Ser Pro Tyr Thr Pro Glu His Ala Ala Ser Ala Pro Thr His Ser  
85 90 95  
Pro Tyr Ala Gln Pro Ser Ser Thr Phe Asp Thr Met Ser Pro Ala Pro  
100 105 110

Val Ile Pro Ser Asn Thr Asp Tyr Pro Gly Pro  
115 120

(2) INFORMATION FOR SEQ ID NO: 11:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 559 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(11) MOLECULE TYPE: cDNA

- (v1) ORIGINAL SOURCE:  
(A) ORGANISM: Homo sapiens

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 11:

GGACCTTCCC	CAGTCAAGCC	GGGGGAATAA	TGAGGTGGTG	GGCGGAACGG	ATTCCAGCAT	60
GGACGTCTTC	CACCTGGAGG	GCATGACTAC	ATCTGTTCAT	CATCCTCGGC	TGCTGCTCA	120
CTAGCTGGGG	AGCCTCTCCC	GCTCGGTCCA	CGCTGCCGGG	CGGCCACGAC	CGTGACCCCT	180
CCCTCGGGG	CGCCAGATC	CATGCCCTGT	CCCACGGGAC	ACCAGTTCCC	TGGGTGTGC	240
AGACCCCGG	GGGCTACCA	TGCTGTACGT	CGGTGACCCC	GCACGGCACC	TGCCACGGC	300
CCAGTTCAT	CTGCTGAGCA	GCACCATGGA	CCAGATGAGC	AGCGCGCGGG	CCTCGGCCAG	360
CCCTACACC	CCAGAGCAGG	CGGCCAGCGT	GCCACCCAC	TGCGCTACG	CACAACCCAG	420
CTCCACCTTC	GACACCATGT	CGCCGGCGCC	TGTCATCCCC	TCCAACACCG	ACTACCCCGG	480
ACCCACCCAC	TTTGAGGTCA	CTTTCCAGCA	GTCCAGCACG	GCCAAGTCAG	CCACCTGGAC	540
GTACTCCCGG	CTCTTGAAG					559

(2) INFORMATION FOR SEQ ID NO: 12:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1764 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(11) MOLECULE TYPE: cDNA

- (v1) ORIGINAL SOURCE:  
(A) ORGANISM: Homo sapiens

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 12:

ATGCTGTACG	TGGGTGACCC	CGCACGGCAC	CTCGCCACGG	CCCAGTTCAA	TCTGCTGAGC	60
AGCACCATGG	ACCAGATGAG	CAGCGCGCGG	GCCTCGGCCA	GCCCTACAC	CCCAGAGCAC	120
GCGGCAGCG	TGCCACCCCA	CTCGCCCTAC	GCACAACCCA	GCTCCACCTT	CGACACCATG	180
TGCGCGGCG	CTGTCTATCC	CTCCAACACC	GACTACCCCG	GACCCACCCA	CTTTGAGGTC	240
ACTTTCCAGC	AGTCCAGCAC	GGCCAAGTCA	GCCACCTGGA	CGTACTCCCC	GCTCTTGAAG	300
AAACTCTACT	GCCAGATGCG	CAAGACATGC	CCCATCCAGA	TCAAGGTGTC	CACCCCGCCA	360
CCCCCAGGCA	CTGCCATCCG	GGCCATGGCT	GTTTACAAGA	AAGCGGAGCA	CGTGACCGAC	420
GTCTGTAAGC	GCTGCCCCAA	CCACGAGCTC	GGGAGGGAAT	TCAACGAAGG	ACAGTCTGCT	480

CCAGCCAGCC ACCTCATCCG CGTGGGAAGGC AATAATCTCT CGCAGTATGT GGATGACCCT 540  
 GTCACCGGCA GGCAGAGCGT CGTGGTGCCC TATGAGCCAC CACAGGTGGG GACCGAATTC 600  
 ACCACCATCC TGTACAACCT CATGTGTAAC AGCAGCTGTG TAGGGGGCAT GAACCGGCGG 660  
 CCGATCCTCA TCATCATCAC CCGGAGATG CGGGATGGGC AGGTGCTGGG CCGCGGCTCC 720  
 TTTGAGGGCC GCATCTGGCC CTGTCTGGC CGCGACCGAA AAGCTGATGA GGACCACTAC 780  
 CGGGAGCAGC AGGCCCTGAA CGAGAGCTCC GCCAAGAAGC GGGCCGCCAG CAAGCGTGCC 840  
 TTCAAGCAGA GCGCCCCCTGC CGTCCCCGCC CTTGGTGCCG GTGTGAAGAA GCGGCGGCAT 900  
 GGAGACGAGG ACACCTACTA CCTTCAGGTG CGAGGCGCGG AGAAGCTTTGA GATCCTGATG 960  
 AAGCTGAAAG AGAGCCTGGA GCTGATGGAG TTGGTGCCGC AGCCACTGGT GGAAGCTCTAT 1020  
 CGGCAGCAGC AGCAGCTCCT ACAGAGGCGC AGTCACCTAC AGCCCCCGTC CTACGGGCGC 1080  
 GTCTCTCGC CCATGAACAA GGTGCACGGG GGCATGAACA AGCTGCCCTC CGTCAACAG 1140  
 CTGGTGGGCC AGCTCCCCC GCACAGTTCG GCAGCTACAC CCAAGCTGGG GCCCGTGGG 1200  
 CCGGGATGC TCAACAACCA TGGCCACGCA GTGCCAGCCA ACGCGAGAT GAGCAGCAGC 1260  
 CACAGCGCCC AGTCCATGGT CTCGGGGTCC CACTGCACTC CGCCACCCCC CTACCACGCC 1320  
 GACCCAGGCC TCGTCAGTTT TTTAACAGGA TTGGGGTGTC CAAACTGCAT CGAGTATTTT 1380  
 ACCTCCCAAG GGTACAGAG CATTTACCAC CTCGAGAACC TGACCATTGA GGAGCTGGGG 1440  
 GCCTTGAGA TCCCCAGCA GTACCGCATG ACCATCTGGC GGGGCTGCA GGAGCTGAAG 1500  
 CAGGGCCAGC ACTACAGCAC CGCGCAGCAG CTGCTCCGCT CTAGCAAGCG GGCCACCATC 1560  
 TCCATCGGG GCTCAGGGGA ACTGCAGCGC CAGCGGGTCA TGGAGGCCGT GCACTTCCGC 1620  
 GTGGGCCACA CCATACCAT CCCCACCGC GCGGGGCCAG GCGCGGCCCC TGACGAGTGG 1680  
 GCGGACTTCG GCTTCGACCT GCCCGACTGC AAGGCCCGCA AGCAGCCCAT CAAGGAGGAG 1740  
 TTCACGGAGG CCGAGATCCA CTGA 1764

## (2) INFORMATION FOR SEQ ID NO: 13:

- (1) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 587 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear

(11) MOLECULE TYPE: protein

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 13:

Met Leu Tyr Val Gly Asp Pro Ala Arg His Leu Ala Thr Ala Gln Phe  
 1 5 10 15  
 Asn Leu Leu Ser Ser Thr Met Asp Gln Met Ser Ser Arg Ala Ala Ser  
 20 25 30  
 Ala Ser Pro Tyr Thr Thr Pro Glu His Ala Ala Ser Val Thr His Ser  
 35 40 45  
 Pro Tyr Ala Gln Pro Ser Ser Thr Phe Asp Thr Met Ser Pro Ala Pro  
 50 55 60  
 Val Ile Pro Ser Asn Thr Asp Tyr Pro Gly Pro His His Phe Glu Val  
 65 70 75 80

Thr Phe Gln Gln Ser Ser Thr Ala Lys Ser Ala Thr Trp Thr Tyr Ser  
 85 90 95  
 Pro Leu Leu Lys Lys Leu Tyr Cys Gln Ile Ala Lys Thr Cys Pro Ile  
 100 105 110  
 Gln Ile Lys Val Ser Thr Pro Pro Pro Gly Thr Ala Ile Arg Ala  
 115 120 125  
 Met Pro Val Tyr Lys Lys Ala Glu His Val Thr Asp Val Val Lys Arg  
 130 135 140  
 Cys Pro Asn His Glu Leu Gly Arg Asp Phe Asn Glu Gly Gln Ser Ala  
 145 150 155 160  
 Pro Ala Ser His Leu Ile Arg Val Glu Gly Asn Asn Leu Ser Gln Tyr  
 165 170 175  
 Val Asp Asp Pro Val Thr Gly Arg Gln Ser Val Val Val Pro Tyr Glu  
 180 185 190  
 Pro Pro Gln Val Gly Thr Glu Phe Thr Thr Ile Leu Tyr Asn Phe Met  
 195 200 205  
 Cys Asn Ser Ser Cys Val Gly Gly Met Asn Arg Arg Pro Ile Leu Ile  
 210 215 220  
 Ile Ile Thr Leu Glu Met Arg Asp Gly Gln Val Leu Gly Arg Arg Ser  
 225 230 235 240  
 Phe Glu Gly Arg Ile Cys Ala Cys Pro Gly Arg Asp Arg Lys Ala Asp  
 245 250 255  
 Glu Asp His Tyr Arg Glu Gln Gln Ala Leu Asn Glu Ser Ser Ala Lys  
 260 265 270  
 Asn Gly Ala Ala Ser Lys Arg Ala Phe Lys Gln Ser Pro Pro Ala Val  
 275 280 285  
 Pro Ala Leu Gly Ala Gly Val Lys Lys Arg Arg His Gly Asp Glu Asp  
 290 295 300  
 Thr Tyr Tyr Leu Gln Val Arg Gly Arg Glu Asn Phe Glu Ile Leu Met  
 305 310 315 320  
 Lys Leu Lys Glu Ser Leu Glu Leu Met Glu Leu Val Pro Gln Pro Leu  
 325 330 335  
 Val Asp Ser Tyr Arg Gln Gln Gln Gln Leu Leu Gln Arg Pro Ser His  
 340 345 350  
 Leu Gln Pro Pro Ser Tyr Gly Pro Val Leu Ser Pro Met Asn Lys Val  
 355 360 365  
 His Gly Gly Met Asn Lys Leu Pro Ser Val Asn Gln Leu Val Gly Gln  
 370 375 380  
 Pro Pro Pro His Ser Ser Ala Ala Thr Pro Asn Leu Gly Pro Val Gly  
 385 390 395 400  
 Pro Gly Met Leu Asn Asn His Gly His Ala Val Pro Ala Asn Gly Glu  
 405 410 415  
 Met Ser Ser Ser His Ser Ala Gln Ser Met Val Ser Gly Ser His Cys  
 420 425 430  
 Thr Pro Pro Pro Tyr His Ala Asp Pro Ser Leu Val Ser Phe Leu  
 435 440 445  
 Thr Gly Leu Gly Cys Pro Asn Cys Ile Glu Tyr Phe Thr Ser Gln Gly  
 450 455 460

Leu Gln Ser Ile Tyr His Leu Gln Asn Leu Thr Ile Glu Asp Leu Gly  
 465 470 475 480  
 Ala Leu Lys Ile Pro Glu Gln Tyr Arg Met Thr Ile Trp Arg Gly Leu  
 485 490 495  
 Gln Asp Leu Lys Gln Gly His Asp Tyr Ser Thr Ala Gln Gln Leu Leu  
 500 505 510  
 Arg Ser Ser Asn Ala Ala Thr Ile Ser Ile Gly Gly Ser Gly Gly Leu  
 515 520 525  
 Gln Arg Gln Arg Val Met Glu Ala Val His Phe Arg Val Arg His Thr  
 530 535 540  
 Ile Thr Ile Pro Asn Arg Gly Gly Pro Gly Gly Pro Asp Glu Trp  
 545 550 555 560  
 Ala Asp Phe Gly Phe Asp Leu Pro Asp Cys Lys Ala Arg Lys Gln Pro  
 565 570 575  
 Ile Lys Glu Glu Phe Thr Glu Ala Glu Ile His  
 580 585

## (2) INFORMATION FOR SEQ ID NO: 14:

- (1) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1521 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(11) MOLECULE TYPE: cDNA

(16) ORIGINAL SOURCE:

(A) ORGANISM: Homo sapiens

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 14:

ATGCTGTACG TCGGTGACCC CGCAGCGGCAC CTCGCCACGG CCCAGTTCAA TGTGTGAGG	60
AGCACCATTGG ACCAGATGAG CAGCCGGCGG GCTCGGGCA GCCCCACAC CCCAGAGCAC	120
GCCGCCACGG TGCCACCCCA CTCGCCCTAC GCACAACCCA GCTCCACCTT CGACACCATTG	180
TCGCCGGCGG CTGTCAATCC CTCGAACACC GACTACCCCG GACCCACCA CTTTGAGGTC	240
ACTTTCCAGC AGTCCAGCAC GGCCAAGTCA GCCACCTGGA CGTACTCCCC GCTTTTGAAG	300
AAACTCTACT GCCAGATCGC CAAGACATGC CCCATCCAGA TCAAGGTGTC CACCCCGCCA	360
CCCCCAGGCA CTCGCATCCG GGCCATGCCT GTTTACAAGA AAGCGGAGCA CGTGACCGAC	420
GTGCTGAAC GCTGCCCAAC CCACGAGCTC GGGAGGGACT TCAACGAAG ACAGTCTGCT	480
CACGCCAGCC ACCTCATCCG CGTGGAGGCG AATAATCTCT CGCAGTATGT GGATGACCCCT	540
GTACCCGGCA GGCAGAGCGT CGTGGTGCCC TATGAGCCAC CACAGGTGGG GACCGAATTC	600
ACCACCATCC TGTACAATT CATGTGTAA AGCAGCTGTG TAGGGGGCAT GAACCCGGCGG	660
CCCATCTCCA TCATCATCAC CCTGGAGATG CGGGATGGGC AGGTGCTGGG CCGCCGGTCC	720
TTTGAGGGCC GCATCTGGCG CTGTCTCTGG CGCAGCCGAA AAGCTGATGA GACCCACTAC	780
CGGGAGCAGC AGGCCCTGAA CGAGAGCTCC GCCAAGAAGC GGGCCGCCAG CAAGCGTGCC	840
TTCAAGCAGA GCCCCCTTGC CTTCCCCGCC CTTGGTGCCG GTGTGAAGAA CGCGCGGCAT	900
GGAGACGAGG ACACGTACTA CTTTCAGGTG CGAGGCGGGG AGAACTTTGA GATCCTGATG	960

```

AAGCTGAAAG AGAGCCTGGA GCTGATGGAG TTGGTGGCGC AGCCACTGGT GGACTCCTAT      1020
CGGCAGCAGC AGCAGCTCCT ACAGAGGCGG CCGCGGGATG CTCAACAACC ATGGCCACGC      1080
AGTGCCAGCC AACGGCGAGA TGAGCAGCAG CCACAGCGCC CAGTCCATGG TCTCGGGGTC      1140
CCACTGCACT CCGCCACCCC CCTACCACGC CGACCCACAG CTCGTGAGGA CCTGGGGGCC      1200
CTGAAGATCC CCGAGCAGTA CCGCATGACC ATCTGGCGGG GCCTGCAGGA CCTGAAGCAG      1260
GGCCAGGACT ACAGACCCGC GCAGCAGCTG CTCGCTCTTA GCAACGCGGC CACCATCTCC      1320
ATCGCGCGGT CAGGGGAAGT GCAGCGCCAG CGGGTCAATG AGGCCGTGCA CTTCGCGGTG      1380
CGCCAGACCA TCACCATCCC CAACCGCGGC GCGCCAGGCG GCGGCCCTGA CGAGTGGGCG      1440
GACTTCGGCT TCGACCTGCC CGACTGCAAG GCCCGCAAGC AGCCCATCAA GGAGGAGTTC      1500
ACGGAGGCGC AGATCCACTG A                                          1521

```

## (2) INFORMATION FOR SEQ ID NO: 15:

- (1) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 506 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear

## (11) MOLECULE TYPE: protein

## (X1) SEQUENCE DESCRIPTION: SEQ ID NO: 15:

```

Met Leu Tyr Val Gly Asp Pro Ala Arg His Leu Ala Thr Ala Gln Phe
1           5           10           15
Asn Leu Leu Ser Ser Thr Met Asp Gln Met Ser Ser Arg Ala Ala Ser
20          25          30
Ala Ser Pro Tyr Thr Pro Glu His Ala Ala Ser Val Pro Thr His Ser
35          40          45
Pro Tyr Ala Gln Pro Ser Ser Thr Phe Asp Thr Met Ser Pro Ala Pro
50          55          60
Val Ile Pro Ser Asn Thr Asp Tyr Pro Gly Pro His His Phe Glu Val
65          70          75          80
Thr Phe Gln Gln Ser Ser Thr Ala Lys Ser Ala Thr Trp Thr Tyr Ser
85          90          95
Pro Leu Leu Lys Lys Leu Tyr Cys Gln Ile Ala Lys Thr Cys Pro Ile
100         105         110
Gln Ile Lys Val Ser Thr Pro Pro Pro Gly Thr Ala Ile Arg Ala
115         120         125
Met Pro Val Tyr Lys Lys Ala Glu His Val Thr Asp Val Val Lys Arg
130         135         140
Cys Pro Asn His Glu Leu Gly Arg Asp Phe Asn Glu Gly Gln Ser Ala
145         150         155         160
Pro Ala Ser His Leu Ile Arg Val Glu Gly Asn Asn Leu Ser Gln Tyr
165         170         175
Val Asp Asp Pro Val Thr Gly Arg Gln Ser Val Val Val Pro Tyr Glu
180         185         190
Pro Pro Gln Val Gly Thr Glu Phe Thr Thr Ile Leu Tyr Asn Phe Met
195         200         205

```

```

Cys Asn Ser Ser Cys Val Gly Gly Met Asn Arg Arg Pro Ile Leu Ile
  210                215                220
Ile Ile Thr Leu Glu Met Arg Asp Gly Gln Val Leu Gly Arg Arg Ser
  225                230                235                240
Phe Glu Gly Arg Ile Cys Ala Cys Pro Gly Arg Asp Arg Lys Ala Asp
                245                250                255
Glu Asp His Tyr Arg Glu Gln Gln Ala Leu Asn Glu Ser Ser Ala Lys
  260                265                270
Asn Gly Ala Ala Ser Lys Arg Ala Phe Lys Gln Ser Pro Pro Ala Val
  275                280                285
Pro Ala Leu Gly Ala Gly Val Lys Lys Arg Arg His Gly Asp Glu Asp
  290                295                300
Thr Tyr Tyr Leu Gln Val Arg Gly Arg Glu Asn Phe Glu Ile Leu Met
  305                310                315                320
Lys Leu Lys Glu Ser Leu Glu Leu Met Glu Leu Val Pro Gln Pro Leu
                325                330                335
Val Asp Ser Tyr Arg Gln Gln Gln Gln Leu Leu Gln Arg Pro Pro Arg
  340                345                350
Asp Ala Gln Gln Pro Trp Pro Arg Ser Ala Ser Gln Arg Arg Asp Glu
  355                360                365
Gln Gln Pro Gln Arg Pro Val His Gly Leu Gly Val Pro Leu His Ser
  370                375                380
Ala Thr Pro Leu Pro Arg Arg Pro Gln Pro Arg Gln Asp Leu Gly Ala
  385                390                395                400
Leu Lys Ile Pro Glu Gln Tyr Arg Met Thr Ile Trp Arg Gly Leu Gln
                405                410                415
Asp Leu Lys Gln Gly His Asp Tyr Ser Thr Ala Gln Gln Leu Leu Arg
  420                425                430
Ser Ser Asn Ala Ala Thr Ile Ser Ile Gly Gly Ser Gly Glu Leu Gln
  435                440                445
Arg Gln Arg Val Met Glu Ala Val His Phe Arg Val Arg His Thr Ile
  450                455                460
Thr Ile Pro Asn Arg Gly Gly Pro Gly Gly Gly Pro Asp Glu Trp Ala
  465                470                475                480
Asp Phe Gly Phe Asp Leu Pro Asp Cys Lys Ala Arg Lys Gln Pro Ile
                485                490                495
Lys Glu Glu Phe Thr Glu Ala Glu Ile His
  500                505

```

## (2) INFORMATION FOR SEQ ID NO: 16:

- (1) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 1870 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear
- (11) MOLECULE TYPE: cDNA
- (VI) ORIGINAL SOURCE:
  - (A) ORGANISM: Homo sapiens
- (IX) FEATURE:
  - (A) NAME/KEY: CDS

(B) LOCATION: 104..1867

(X1) SEQUENCE DESCRIPTION: SEQ ID NO: 16:

TSCCCGGGGG	TGCGACGGCT	GCACGGGAACC	AGACAGCACC	TACTTCGACC	TTCCCCAGTC	60
AAGCCSGGGG	AATAATGAGG	TGGTGGGGGG	AACGGATTCC	AGC	ATG GAC GTC TTC	115
				Met	Asp Val Phe	
				1		
CAC CTG GAG GGC	ATG ACT ACA TCT	GTC ATG GCC	CAG TTC AAT	CTG CTG		163
His Leu Glu Gly	Met Thr Thr Ser Val	Met Ala Gln	Phe Asn Leu Leu			
5	10	15	20			
AGC AGC ACC ATG	GAC GAG ATG AGC	AGC CGC GCG	GCC TCG GCC	AGC CCC		211
Ser Ser Thr Met	Asp Gln Met Ser	Ser Ser Arg	Ala Ala Ser	Ala Ser Pro		
	25	30	35			
TAC ACC CCA GAG	CAC GCC GCC AGC	GTG CCC ACC	CAC TCG CCC	TAC GCA		259
Tyr Thr Pro Glu	His Ala Ala Ser	Val Thr Thr	His Ser	Tyr Ala		
	40	45	50			
CAA CCC AGC TCC	ACC TTC GAC ACC	ATG TCG CCG	GCG CCT GTC	ATC CCC		307
Gln Pro Ser Ser	Thr Phe Asp Thr	Met Ser Pro	Ala Pro Val	Ile Pro		
	55	60	65			
TCC AAC ACC GAC	TAC CCC GGA CCC	CAC CAC TTT	GAG GTC ACT	TTC CAG		355
Ser Asn Thr Asp	Tyr Pro Gly Pro	His His Phe	Glu Val Thr	Phe Gln		
	70	75	80			
CAG TCC AGC ACG	GCC AAG TCA GCC	ACC TGG ACG	TAC TCC CCG	CTC TTG		403
Gln Ser Ser Thr	Ala Lys Ser Ala	Thr Trp Thr	Tyr Ser Pro	Leu Leu		
	85	90	95	100		
AAG AAA CTC TAC	TGC CAG ATC GCC	AAG ACA TGC	CCC ATC CAG	ATC AAG		451
Lys Lys Leu Tyr	Cys Gln Ile Ala	Lys Thr Cys	Pro Ile Gln	Ile Lys		
	105	110	115			
GTG TCC ACC CCG	CCA CCC CCA GGC	ACT GCC ATC	CGG GCC ATG	CCT GTT		499
Val Ser Thr Pro	Pro Pro Pro Gly	Thr Ala Ile	Arg Ala Met	Pro Val		
	120	125	130			
TAC AAG AAA GCG	GAG CAC GTG ACC	GAC GTC GTG	AAA CGC TGC	CCC AAC		547
Tyr Lys Lys Ala	Glu His Val Thr	Asp Val Val	Lys Arg Cys	Pro Asn		
	135	140	145			
CAC GAG CTC GGG	AGG GAC TTC AAC	GAA GGA CAG	TCT GCT CCA	GCC AGC		595
His Glu Leu Gly	Arg Asp Phe	Asn Asn Glu	Gly Ser Ala	Pro Ala Ser		
	150	155	160			
CAC CTC ATC CGC	GTG GAA GGC	AAT AAT CTC	TCG CAG TAT	GTG GAT		643
His Leu Ile Arg	Val Glu Gly	Asn Asn Leu	Ser Gln Tyr	Val Asp Asp		
	165	170	175	180		
CCT GTC ACC GGC	AGG CAG AGC	GTC GTG GTG	CCC TAT GAG	CCA CCA		691
Pro Val Thr Gly	Arg Gln Ser	Val Val Val	Pro Tyr Gln	Pro Gln		
	185	190	195			
GTG GGG ACG GAA	TTC ACC ACC	ATC CTG TAC	AAC TTC ATG	TGT AAC		739
Val Gly Thr Glu	Phe Thr Thr	Ile Leu Tyr	Asn Phe Met	Cys Asn Ser		
	200	205	210			
AGC TGT GTA GGG	GGC ATG AAC	CGG CGG CCC	ATC CTC ATC	ATC ACC		787
Ser Cys Val Gly	Gly Met Asn	Arg Arg Pro	Ile Leu Ile	Ile Thr		
	215	220	225			
CTG GAG ATG CGG	GAT GGG CAG	GTG CTG GGC	CGC CGG TCC	TTT GAG		835
Leu Glu Met Arg	Asp Gly Gln	Val Leu Gly	Arg Arg Ser	Phe Glu Gly		
	230	235	240			
CGC ATC TGC GCC	TGT CCT GGC	CGC GAC CGA	AAA GCT GAT	GAG GAC		883

Arg Ile Cys Ala Cys Pro Gly Arg Asp Arg Lys Ala Asp Glu Asp His 245 250 255 260	
TAC CGG GAG CAG CAG GCC CTG AAC GAG AGC TCC GCC AAG AAC GGG GCC Tyr Arg Glu Gln Gln Ala Leu Asn Glu Ser Ser Ala Lys Asn Gly Ala 265 270 275	931
GCC ACC AAG CGT GCC TTC AAG CAG AGC CCC CCT GCC GTC CCC GCC CTT Ala Ser Lys Arg Ala Phe Lys Gln Ser Pro Pro Ala Val Pro Ala Leu 280 285 290	979
GGT GCC GGT GTG AAG AAG CGG CGG CAT GGA GAC GAG GAC ACG TAC TAC Gly Ala Gly Val Lys Lys Arg Arg His Gly Asp Thr Tyr Tyr 295 300 305	1027
CTT CAG GTG CGA GGC CGG GAG AAC TTT GAG ATC CTG ATG AAG CTG AAA Leu Gln Val Arg Gly Arg Glu Asn Phe Glu Ile Leu Met Lys Leu Lys 310 315 320	1075
GAG AGC CTG GAG CTG ATG GAG TTG GTG CCG CAG CCA CTG GTG GAC TCC Glu Ser Leu Glu Leu Met Glu Leu Val Pro Gln Pro Leu Val Asp Ser 325 330 335 340	1123
TAT CGG CAG CAG CAG CAG CTC CTA CAG AGG CCG AGT CAC CTA CAG CCC Tyr Arg Gln Gln Gln Gln Leu Leu Gln Arg Pro Ser His Leu Gln Pro 345 350 355	1171
CCG TCC TAC GGG CCG GTC CTC TCG CCC ATG AAC AAG GTG CAC GGG GGC Pro Ser Tyr Gly Pro Val Leu Ser Pro Met Asn Lys Val His Gly Gly 360 365 370	1219
ATG AAC AAG CTG CCC TCC GTC AAC CAG CTG GTG GGC CAG CCT CCC CCG Met Asn Lys Leu Pro Ser Val Asn Gln Leu Val Gly Gln Pro Pro Pro 375 380 385	1267
CAC AGT TCG GCA GCT ACA CCC AAC CTG GGG CCC GTG GGC CCC GGG ATG His Ser Ser Ala Ala Thr Pro Asn Leu Gly Pro Val Gly Pro Gly Met 390 395 400	1315
CTC AAC AAC CAT GGC CAC GCA GTG CCA GCC AAC GGC GAG ATG AGC AGC Leu Asn Asn His Gly His Ala Val Pro Ala Asn Gly Glu Met Ser Ser 405 410 415 420	1363
AGC CAC AGC GCC CAG TCC ATG GTC TCG GGG TCC CAC TGC ACT CCG CCA Ser His Ser Ala Gln Ser Met Val Ser Gly Ser His Cys Thr Pro Pro 425 430 435	1411
CCC CCC TAC CAC GCC GAC CCC AGC CTC GTC AGT TTT TTA ACA GGA TTG Pro Pro Tyr His Ala Asp Pro Ser Leu Val Ser Phe Leu Thr Gly Leu 440 445 450	1459
GGG TGT CCA AAC TGC ATC GAG TAT TTC ACC TCC CAA GGG TTA CAG AGC Gly Cys Pro Asn Cys Ile Glu Tyr Phe Thr Ser Gln Gly Leu Gln Ser 455 460 465	1507
ATT TAC CAC CTG CAG AAC CTG ACC ATT GAG GAC CTG GGG GCC CTG AAG Ile Tyr His Leu Gln Asn Leu Thr Ile Glu Asp Leu Gly Ala Leu Lys 470 475 480	1555
ATC CCC GAG CAG TAC CGC ATG ACC ATC TGG CCG GGC CTG CAG GAC CTG Ile Pro Glu Gln Tyr Arg Met Thr Ile Trp Arg Gly Leu Gln Asp Leu 485 490 495 500	1603
AAG CAG GGC CAC GAC TAC AGC ACC GCG CAG CAG CTG CTC CGC TCT AGC Lys Gln Gly His Asp Tyr Ser Thr Ala Gln Gln Leu Leu Arg Ser Ser 505 510 515	1651
AAC GCG GCC ACC ATC TCC ATC GGC GGC TCA GGG GAA CTG CAG CGC CAG Asn Ala Ala Thr Ile Ser Ile Gly Gly Ser Gly Glu Leu Gln Arg Gln 520 525 530	1699
CGG GTC ATG GAG GCC GTG CAC TTC CGC GTG CGC CAC ACC ATC ACC ATC	1747

Arg Val Met Glu Ala Val His Phe Arg Val Arg His Thr Ile Thr Ile  
 535 540 545

CCC AAC CGC GGC GGC CCA GGC GGC GGC CCT GAC GAG TGG GCG GAC TTC 1795  
 Pro Asn Arg Gly Gly Pro Gly Gly Gly Pro Asp Glu Trp Ala Asp Phe  
 550 555 560

GGC TTC GAC CTG CCC GAC TGC AAG GCC CGC AAG CAG CCC ATC AAG GAG 1843  
 Gly Phe Asp Leu Pro Asp Cys Lys Ala Arg Lys Gln Pro Ile Lys Glu  
 565 570 575 580

GAG TTC ACG GAG GCC GAG ATC CAC TGA 1870  
 Glu Phe Thr Glu Ala Glu Ile His  
 585

## (2) INFORMATION FOR SEQ ID NO: 17:

- (1) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 588 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear

(11) MOLECULE TYPE: protein

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 17:

Met Asp Val Phe His Leu Glu Gly Met Thr Thr Ser Val Met Ala Gln  
 1 5 10 15

Phe Asn Leu Leu Ser Ser Thr Met Asp Gln Met Ser Ser Arg Ala Ala  
 20 25 30

Ser Ala Ser Pro Tyr Thr Pro Glu His Ala Ala Ser Val Pro Thr His  
 35 40 45

Ser Pro Tyr Ala Gln Pro Ser Ser Thr Phe Asp Thr Met Ser Pro Ala  
 50 55 60

Pro Val Ile Pro Ser Asn Thr Asp Tyr Pro Gly Pro His His Phe Glu  
 65 70 75 80

Val Thr Phe Gln Gln Ser Ser Thr Ala Lys Ser Ala Thr Trp Thr Tyr  
 85 90 95

Ser Pro Leu Leu Lys Lys Leu Tyr Cys Gln Ile Ala Lys Thr Cys Pro  
 100 105 110

Ile Gln Ile Lys Val Ser Thr Pro Pro Pro Gly Thr Ala Ile Arg  
 115 120 125

Ala Met Pro Val Tyr Lys Lys Ala Glu His Val Thr Asp Val Val Lys  
 130 135 140

Arg Cys Pro Asn His Glu Leu Gly Arg Asp Phe Asn Glu Gly Gln Ser  
 145 150 155 160

Ala Pro Ala Ser His Leu Ile Arg Val Glu Gly Asn Asn Leu Ser Gln  
 165 170 175

Tyr Val Asp Asp Pro Val Thr Gly Arg Gln Ser Val Val Val Pro Tyr  
 180 185 190

Glu Pro Pro Gln Val Gly Thr Glu Phe Thr Thr Ile Leu Tyr Asn Phe  
 195 200 205

Met Cys Asn Ser Ser Cys Val Gly Gly Met Asn Arg Arg Pro Ile Leu  
 210 215 220

Ile Ile Ile Thr Leu Glu Met Arg Asp Gly Gln Val Leu Gly Arg Arg  
 225 230 235 240

Ser Phe Glu Gly Arg Ile Cys Ala Cys Pro Gly Arg Asp Arg Lys Ala  
 245 250 255  
 Asp Glu Asp His Tyr Arg Glu Gln Gln Ala Leu Asn Glu Ser Ser Ala  
 260 265 270  
 Lys Asn Gly Ala Ala Ser Lys Arg Ala Phe Lys Gln Ser Pro Pro Ala  
 275 280 285  
 Val Pro Ala Leu Gly Ala Gly Val Lys Lys Arg Arg His Gly Asp Glu  
 290 295 300  
 Asp Thr Tyr Tyr Leu Gln Val Arg Gly Arg Glu Asn Phe Glu Ile Leu  
 305 310 315 320  
 Met Lys Leu Lys Glu Ser Leu Glu Leu Met Glu Leu Val Pro Gln Pro  
 325 330 335  
 Leu Val Asp Ser Ser Tyr Arg Gln Gln Gln Leu Leu Gln Arg Pro Ser  
 340 345 350  
 His Leu Gln Pro Pro Ser Tyr Gly Pro Val Leu Ser Pro Met Asn Lys  
 355 360 365  
 Val His Gly Gly Met Asn Lys Leu Pro Ser Val Asn Gln Leu Val Gly  
 370 375 380  
 Gln Pro Pro Pro His Ser Ser Ala Ala Thr Pro Asn Leu Gly Pro Val  
 385 390 395 400  
 Gly Pro Gly Met Leu Asn Asn His Gly His Ala Val Pro Ala Asn Gly  
 405 410 415  
 Glu Met Ser Ser Ser His Ser Ala Gln Ser Met Val Ser Gly Ser His  
 420 425 430  
 Cys Thr Pro Pro Pro Tyr His Ala Asp Pro Ser Leu Val Ser Phe  
 435 440 445  
 Leu Thr Gly Leu Gly Cys Pro Asn Cys Ile Glu Tyr Phe Thr Ser Gln  
 450 455 460  
 Gly Leu Gln Ser Ile Tyr His Leu Gln Asn Leu Thr Ile Glu Asp Leu  
 465 470 475 480  
 Gly Ala Leu Lys Ile Pro Glu Gln Tyr Arg Met Thr Ile Trp Arg Gly  
 485 490 495  
 Leu Gln Asp Leu Lys Gln Gly His Asp Tyr Ser Thr Ala Gln Gln Leu  
 500 505 510  
 Leu Arg Ser Ser Asn Ala Ala Thr Ile Ser Ile Gly Gly Ser Gly Glu  
 515 520 525  
 Leu Gln Arg Gln Arg Val Met Glu Ala Val His Phe Arg Val Arg His  
 530 535 540  
 Thr Ile Thr Ile Pro Asn Arg Gly Gly Pro Gly Gly Gly Pro Asp Glu  
 545 550 555 560  
 Trp Ala Asp Phe Gly Phe Asp Leu Pro Asp Cys Lys Ala Arg Lys Gln  
 565 570 575  
 Pro Ile Lys Glu Glu Phe Thr Glu Ala Glu Ile His  
 580 585

(2) INFORMATION FOR SEQ ID NO: 18:

- (1) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1817 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(11) MOLECULE TYPE: cDNA

(v1) ORIGINAL SOURCE:

(A) ORGANISM: Homo sapiens

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 18:

ATGGCCAGT CCACCCGAC CTCCCCTGAT GGGGGCACCA CGTTTGAGCA CCTCTGGAGC	60
TCTCTGGAAC CAGACAGCAC CTACTTCGAC CTTCCCCAGT CAAGCCGGGG GAATAATGAG	120
GTGGTGGGCG GAACGGATTG CAGCATGGAC GTCTTCACCC TGGAGGGCAT GACTACATCT	180
GTGATGGCCC AGTTCAATCT GCTGAGCAGC ACCATGGACC AGATGAGCAG CGGCGCGGGC	240
TCGGCCAGCC CTTACACCCC AGAGCAGGCC GCCAGCGTGC CCACCCACTC GCCCTACGCA	300
CAACCCAGCT CCACCTTCGA CACCATGTGC CCGGCGCGTG TCATCCCCCT CAACCCGAC	360
TACCCCGGAC CCCACCACTT TGAGGTCACG TTCACGAGT CCAGCAGCGC CAAGTCAGCC	420
ACCTGGAGCT ACTCCCCGCT CTTGAAGAAA CTCTACTGCC AGATCGCCAA GACATGCCCC	480
ATCCAGATCA AGGTGTCCAC CCGGCCACCC CCAGGCACTG CCATCCGGGG CATGCCTGTT	540
TACAAGAAAG CGGAGCAGCT GACCGACGTC GTGAAACGCT GCCCCAAACA CGAGCTCGGG	600
AGGGACTTCA ACGAAGGACA GTCTGCTCCA GCCAGCCACC TCATCCGGCT GGAAGGCAAT	660
AATCTCTGCG AGTATGTGGA TGACCTGTGC ACCGGCAGGC AGAGCGTCTG GTGCGCCTAT	720
GAGCCACCAC AGGTGGGGAC GGAATTCACC ACCATCCTGT ACAACTTCAT GTGTAACAGC	780
AGCTGTGTAG GGGGCGATGA CCGGCGGCCC ATCCTCATCA TCATCACCTC GGAGATGCGG	840
GATGGGCAGG TGCTGGGCGC CGGTCTCTTT GAGGGCCGCA TCTGCGCCTG TCCTGCGCGC	900
GACCGAAAAG CTGATGAGGA CCACTACCGG GAGCAGCAGG CCCTGAACGA GAGCTCCGCC	960
AAGAACGGGG CCGCCAGCAA GCGTGCTTC AAGCAGAGCC CCCCTGCCGT CCCCGCCCTT	1020
GGTGCCGGTG TGAAGAAGCG GCGGCATGGA GACGAGGACA CGTACTACCT TCAGGTGCGA	1080
GGCCGGGAGA ACTTTGAGAT CCTGATGAAG CTGAAAGAGA GCCTGGAGCT GATGGAGTTG	1140
GTGCCGACG CACTGTGGA CTCTATCGG CAGCAGCAGC AGCTCTTACA GAGGCCGAGT	1200
CACCTACAGC CCCCGCTCTA CGGGCCGGTC CTCTCGCCCA TGAACAAGGT GCACGGGGGC	1260
ATGAACAAGC TGCCCTCCGT CAACCAAGCTG GTGGGCCAGC CTCGCCCGCA CAGTTCGGCA	1320
GCTACACCCA ACCTGGGGCC CGTGGGCCCC GGGATGTCTA ACAACCATGG CCACGCAGTG	1380
CCAGCCAAAG GCGAGATGAG CAGCAGCCAC AGCGCCCACT CCATGGTCTC GGGGTCCAC	1440
TGCACTCCGC CACCCCTCTA CCAGGCCGAC CCCAGCCTCG TCAGGACCTG GGGGCCCTGA	1500
AGATCCCCGA CGAGTACCGC ATGACCATCT GCGGGGGCCT GCAGGACCTG AAGCAGGGCC	1560
ACGACTACAG CACCGCGCAG CAGCTGCTCC GCTCTAGCAA CGCGGCCACC ATCTCCATCG	1620
GCGGCTCAGG GGAAGTCGAG CGCCAGCGGG TCATGAGGCG CGTGCACTTC CCGGTGCGGC	1680
ACACCATCAC CATCCCCAAC CGCGCGGGCC CAGGCGGGCG CCCTGACGAG TGGCGGGA	1740
TCGGCTTCGA CCTGCCCGAC TGCAAGGCCC GCAAGCAGCC CATCAAGGAG GAGTTCACGG	1800
AGGCCGAGAT CCACTGA	1817

## (2) INFORMATION FOR SEQ ID NO: 19:

- (1) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 499 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear

(11) MOLECULE TYPE: protein

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 19:

```

Met Ala Gln Ser Thr Ala Thr Ser Pro Asp Gly Gly Thr Thr Phe Glu
1      5      10
His Leu Trp Ser Ser Leu Glu Pro Asp Ser Thr Tyr Phe Asp Leu Pro
20     25     30
Gln Ser Ser Arg Gly Asn Asn Glu Val Val Gly Gly Thr Asp Ser Ser
35     40     45
Met Asp Val Phe His Leu Glu Gly Met Thr Thr Ser Val Met Ala Gln
50     55     60
Phe Asn Leu Leu Ser Ser Thr Met Asp Gln Met Ser Ser Arg Ala Ala
65     70     75     80
Ser Ala Ser Pro Tyr Thr Pro Glu His Ala Ala Ser Val Pro Thr His
85     90     95
Ser Pro Tyr Ala Gln Pro Ser Ser Thr Phe Asp Thr Met Ser Pro Ala
100    105    110
Pro Val Ile Pro Ser Asn Thr Asp Tyr Pro Gly Pro His His Phe Glu
115    120    125
Val Thr Phe Gln Gln Ser Ser Thr Ala Lys Ser Ala Thr Trp Thr Tyr
130    135    140
Ser Pro Leu Leu Lys Lys Leu Tyr Cys Gln Ile Ala Lys Thr Cys Pro
145    150    155    160
Ile Gln Ile Lys Val Ser Thr Pro Pro Pro Pro Gly Thr Ala Ile Arg
165    170    175
Ala Met Pro Val Tyr Lys Lys Ala Glu His Val Thr Asp Val Val Lys
180    185    190
Arg Cys Pro Asn His Glu Leu Gly Asp Arg Phe Asn Glu Gly Gln Ser
195    200    205
Ala Pro Ala Ser His Leu Ile Arg Val Glu Gly Asn Asn Leu Ser Gln
210    215    220
Tyr Val Asp Asp Pro Val Thr Gly Arg Gln Ser Val Val Val Pro Tyr
225    230    235    240
Glu Pro Pro Gln Val Gly Thr Glu Phe Thr Thr Ile Leu Tyr Asn Phe
245    250    255
Met Cys Asn Ser Ser Cys Val Gly Gly Met Asn Arg Arg Pro Ile Leu
260    265    270
Ile Ile Ile Thr Leu Glu Met Arg Asp Gly Gln Val Leu Gly Arg Arg
275    280    285
Ser Phe Glu Gly Arg Ile Cys Ala Cys Pro Gly Arg Asp Arg Lys Ala
290    295    300
Asp Glu Asp His Tyr Arg Glu Gln Gln Ala Leu Asn Glu Ser Ser Ala
305    310    315    320

```

[illegible]

(2) INFORMATION FOR SEQ ID NO: 20:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 17 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(11) MOLECULE TYPE: DNA

(111) HYPOTHETICAL: NO

(111) ANTI-SENSE: NO

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 20:

GCGAGCTGCC CTCGGAG

17

## (2) INFORMATION FOR SEQ ID NO: 21:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 19 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(11) MOLECULE TYPE: DNA

(111) ANTI-SENSE: YES

(X1) SEQUENCE DESCRIPTION: SEQ ID NO: 21:

GGTTCTGCAG GTGACTCAG

19

(2) INFORMATION FOR SEQ ID NO: 22:

- (1) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 18 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(11) MOLECULE TYPE: DNA

(111) ANTI-SENSE: NO

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 22:

GCCATGCCTG TCTACAAG

18

(2) INFORMATION FOR SEQ ID NO: 23:

- (1) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 18 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(11) MOLECULE TYPE: DNA

(111) ANTI-SENSE: YES

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 23:

ACCAGCTGGT TGACGGAG

18

(2) INFORMATION FOR SEQ ID NO: 24:

- (1) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 21 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(11) MOLECULE TYPE: DNA

(111) ANTI-SENSE: NO

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 24:

GTCAACCAGC TGGTGGGCCA G

21

(2) INFORMATION FOR SEQ ID NO: 25:

- (1) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 16 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(11) MOLECULE TYPE: DNA

(111) ANTI-SENSE: YES

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 25:

GIGGATCTCG GCCTCC

16

(2) INFORMATION FOR SEQ ID NO: 26:

- (1) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 17 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(11) MOLECULE TYPE: DNA

(111) ANTI-SENSE: NO

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 26:

AGGCCGGCGT GGGGAAG

17

(2) INFORMATION FOR SEQ ID NO: 27:

- (1) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 19 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(11) MOLECULE TYPE: DNA

(111) ANTI-SENSE: YES

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 27:

CTTGGCGATC TGGCAGTAG

19

(2) INFORMATION FOR SEQ ID NO: 28:

- (1) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 17 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(11) MOLECULE TYPE: DNA

(111) ANTI-SENSE: NO

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 28:

GCGGCCACGA CCGTGAC

17

(2) INFORMATION FOR SEQ ID NO: 29:

- (1) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 18 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(11) MOLECULE TYPE: DNA

(111) ANTI-SENSE: YES

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 29:

GGCAGCTTGG GTCTCTGG

18

(2) INFORMATION FOR SEQ ID NO: 30:

- (1) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 19 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(11) MOLECULE TYPE: DNA

(111) ANTI-SENSE: NO

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 30:

CTGTACGTCG GTGACCCC

18

(2) INFORMATION FOR SEQ ID NO: 31:

- (1) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 19 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(11) MOLECULE TYPE: DNA

(111) ANTI-SENSE: YES

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 31:

TCAGTGGATC TCGGCCTC

18

(2) INFORMATION FOR SEQ ID NO: 32:

- (1) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 18 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(11) MOLECULE TYPE: DNA

(111) ANTI-SENSE: NO

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 32:

AGGGGACGCA GCGAAACC

18

(2) INFORMATION FOR SEQ ID NO: 33:

- (1) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 19 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(11) MOLECULE TYPE: DNA

(111) ANTI-SENSE: YES

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 33:

CCATCAGCTC CAGGCTCTC

19

(2) INFORMATION FOR SEQ ID NO: 34:

- (1) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 18 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear
- (11) MOLECULE TYPE: DNA
- (111) ANTI-SENSE: YES

(X1) SEQUENCE DESCRIPTION: SEQ ID NO: 34:

CCAGGACAGG CGCAGATG

18

(2) INFORMATION FOR SEQ ID NO: 35:

- (1) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 19 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear
- (11) MOLECULE TYPE: DNA
- (111) ANTI-SENSE: YES

(X1) SEQUENCE DESCRIPTION: SEQ ID NO: 35:

GATGAGGTGG CTGGCTGGA

19

(2) INFORMATION FOR SEQ ID NO: 36:

- (1) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 19 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear
- (11) MOLECULE TYPE: DNA
- (111) ANTI-SENSE: YES

(X1) SEQUENCE DESCRIPTION: SEQ ID NO: 36:

TGGTCAGGTT CTGCAGGTG

19

(2) INFORMATION FOR SEQ ID NO: 37:

- (1) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 18 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear
- (11) MOLECULE TYPE: DNA
- (111) ANTI-SENSE: NO

(X1) SEQUENCE DESCRIPTION: SEQ ID NO: 37:

CACCTACTCC AGGGATGC

18

(2) INFORMATION FOR SEQ ID NO: 38:

- (1) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 21 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(11) MOLECULE TYPE: DNA

(111) ANTI-SENSE: YES

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 38:

AGGAAAATAG AAGCGTCAGT C

21

(2) INFORMATION FOR SEQ ID NO: 39:

- (1) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 18 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(11) MOLECULE TYPE: DNA

(111) ANTI-SENSE: NO

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 39:

CAGGCCCACT TGCCTGCC

18

(2) INFORMATION FOR SEQ ID NO: 40:

- (1) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 19 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(11) MOLECULE TYPE: DNA

(111) ANTI-SENSE: YES

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 40:

CTGTCOCCAA GCTGATGAG

19

## CLAIMS

1. Purified polypeptide, comprising an amino acid sequence selected from the group consisting of:

- a) the sequence SEQ ID No. 2;
- b) the sequence SEQ ID No. 4;
- c) the sequence SEQ ID No. 6;
- d) the sequence SEQ ID No. 8;
- e) the sequence SEQ ID No. 10;
- f) the sequence SEQ ID No. 13;
- g) the sequence SEQ ID No. 15;
- h) the sequence SEQ ID No. 17;
- i) the sequence SEQ ID No. 19;

and j) any biologically active sequence derived from SEQ ID No. 2, SEQ ID No. 4, SEQ ID No. 6, SEQ ID No. 8, SEQ ID No. 10, SEQ ID No. 13, SEQ ID No. 15, SEQ ID No. 17 or SEQ ID No. 19.

2. Polypeptide according to Claim 1, characterized in that it comprises the amino acid sequence selected from the group consisting of SEQ ID No. 6, SEQ ID No. 13, SEQ ID No. 15, SEQ ID No. 17 and SEQ ID No. 19.

3. Polypeptide according to Claim 1, characterized in that it comprises the sequence lying between:

- residue 110 and residue 310 of SEQ ID No. 2 or 6;
- residue 60 and residue 260 of SEQ ID No. 8.

4. Polypeptide according to Claim 1, characterized in that it results from an alternative splicing of the messenger RNA of the corresponding gene.

5. Polypeptide according to any one of the preceding claims, characterized in that it is a recombinant polypeptide produced in the form of a fusion protein.

6. Isolated nucleic acid sequence coding for a polypeptide according to any one of the preceding claims.

7. Isolated nucleic acid sequence according to Claim 6, characterized in that it is selected from the group consisting of :

- a) the sequence SEQ ID No. 1;
- b) the sequence SEQ ID No. 3;
- c) the sequence SEQ ID No. 5;

- d) the sequence SEQ ID No. 7;  
e) the sequence SEQ ID No. 9;  
f) the sequence SEQ ID No. 11;  
g) the sequence SEQ ID No. 12;  
5 h) the sequence SEQ ID No. 14;  
i) the sequence SEQ ID No. 16;  
j) the sequence SEQ ID No. 18;  
k) the nucleic acid sequences capable of hybridizing specifically with the sequence SEQ ID No. 1, SEQ ID No. 3, SEQ ID No. 5, SEQ ID No. 7, SEQ ID No. 9, SEQ ID No. 11, SEQ ID No. 12, SEQ ID No. 14, SEQ ID No. 16 or SEQ ID No. 18 or with the sequences complementary to them, or of hybridizing specifically with their proximal sequences;  
10 and l) the sequences derived from the sequences a), b), c), d), e), f), g), h), i), j) or k) as a result of the degeneracy of the genetic code, mutation, deletion, insertion, and alternative splicing or an allelic variability.
- 20 8. Nucleotide sequence according to Claim 6, characterized in that it is a sequence selected from SEQ ID No. 5, SEQ ID No. 12, SEQ ID No. 14, SEQ ID No. 16 and SEQ ID No. 18 coding, respectively, for the polypeptide of sequences SEQ ID No. 6, SEQ ID No. 13, SEQ ID No. 15, SEQ ID No. 17 and SEQ ID No. 19.
- 25 9. Cloning and/or expression vector containing a nucleic acid sequence according to any one of Claims 6 to 8.
10. Vector according to Claim 9, characterized in  
30 that it is the plasmid pSE1.
11. Host cell transfected by a vector according to Claim 9 or 10.
12. Transfected host cell according to Claim 11, characterized in that it is *E. coli* MC 1061.
- 35 13. Nucleotide probe or nucleotide primer, characterized in that it hybridizes specifically with any one of the sequences according to Claims 6 to 8 or the sequences complementary to them or the corresponding messenger RNAs or the corresponding  
40 genes.

14. Probe or primer according to Claim 13, characterized in that it contains at least 16 nucleotides.

15. Probe or primer according to Claim 13, characterized in that it comprises the whole of the sequence of the gene coding for one of the polypeptides of Claim 1.

16. Nucleotide probe or primer selected from the group consisting of the following oligonucleotides or sequences complementary to them:

SEQ ID No. 20: GCG AGC TGC CCT CGG AG  
 SEQ ID No. 21: GGT TCT GCA GGT GAC TCA G  
 SEQ ID No. 22: GCC ATG CCT GTC TAC AAG  
 SEQ ID No. 23: ACC AGC TGG TTG ACG GAG  
 SEQ ID No. 24: GTC AAC CAG CTG GTG GGC CAG  
 SEQ ID No. 25: GTG GAT CTC GGC CTC C  
 SEQ ID No. 26: AGG CCG GCG TGG GGA AG  
 SEQ ID No. 27: CTT GGC GAT CTG GCA GTA G  
 SEQ ID No. 28: GCG GCC ACG ACC GTG AC  
 SEQ ID No. 29: GGC AGC TTG GGT CTC TGG  
 SEQ ID No. 30: CTG TAC GTC GGT GAC CCC  
 SEQ ID No. 31: TCA GTG GAT CTC GGC CTC  
 SEQ ID No. 32: AGG GGA CGC AGC GAA ACC  
 SEQ ID No. 33: CCA TCA GCT CCA GGC TCT C  
 SEQ ID No. 34: CCA GGA CAG GCG CAG ATG  
 SEQ ID No. 35: GAT GAG GTG GCT GGC TGG A  
 SEQ ID No. 36: TGG TCA GGT TCT GCA GGT G  
 SEQ ID No. 37: CAC CTA CTC CAG GGA TGC  
 SEQ ID No. 38: AGG AAA ATA GAA GCG TCA GTC  
 SEQ ID No. 39: CAG GCC CAC TTG CCT GCC  
 and SEQ ID No. 40: CTG TCC CCA AGC TGA TGA G

17. Use of a sequence according to any one of Claims 6 to 8, for the manufacture of oligonucleotide primers for sequencing reactions or specific amplification reactions according to the PCR technique or any variant of the latter.

18. Nucleotide primer pair, characterized in that it comprises the primers selected from the group consisting of the following sequences:

- a) sense primer: GCG AGC TGC CCT CGG AG (SEQ ID No. 20)  
antisense primer: GGT TCT GCA GGT GAC TCA G (SEQ ID No. 21)
- b) sense primer: GCC ATG CCT GTC TAC AAG (SEQ ID No. 22)  
5 antisense primer: ACC AGC TGG TTG ACG GAG (SEQ ID No. 23)
- c) sense primer: GTC AAC CAG CTG GTG GGC CAG (SEQ ID No. 24)  
antisense primer: GTG GAT CTC GGC CTC C (SEQ ID No. 25)
- 10 d) sense primer: AGG CCG GCG TGG GGA AG (SEQ ID No. 26)  
antisense primer: CTT GGC GAT CTG GCA GTA G (SEQ ID No. 27)
- e) sense primer: GCG GCC ACG ACC GTG A (SEQ ID No. 28)  
antisense primer: GGC AGC TTG GGT CTC TGG (SEQ ID No. 29)
- 15 f) sense primer: CTG TAC GTC GGT GAC CCC (SEQ ID No. 30)  
antisense primer: TCA GTG GAT CTC GGC CTC (SEQ ID No. 31)
- g) sense primer: AGG GGA CGC AGC GAA ACC (SEQ ID No. 32)  
20 antisense primer: GGC AGC TTG GGT CTC TGG (SEQ ID No. 29)
- h) sense primer: CCCCCCCCCCCCCCN (where N equals G, A or T)  
antisense primer: CCA TCA GCT CCA GGC TCT C (SEQ ID No. 33)
- 25 i) sense primer: CCCCCCCCCCCCCCN (where N equals G, A or T)  
antisense primer: CCA GGA CAG GCG CAG ATG (SEQ ID No. 34)
- j) sense primer: CCCCCCCCCCCCCCN (where N equals G, A or T)  
antisense primer: CTT GGC GAT CTG GCA GTA G (SEQ ID No. 27)
- 30 k) sense primer: CAC CTA CTC CAG GGA TGC (SEQ ID No. 37)  
antisense primer: AGG AAA ATA GAA GCG TCA GTC (SEQ ID No. 38)
- 35 and l) sense primer: CAG GCC CAC TTG CCT GCC (SEQ ID No. 39)  
antisense primer: CTG TCC CCA AGC TGA TGA G (SEQ ID No. 40)
19. Use of a sequence according to any one of Claims 6 to 8, which is usable in gene therapy.
20. Use of a sequence according to any one of  
40 Claims 6 to 8, for the production of diagnostic

nucleotide probes or primers, or of antisense sequences which are usable in gene therapy.

21. Use of nucleotide primers according to any one of Claims 6 to 8, for sequencing.
- 5 22. Use of a probe or primer according to any one of Claims 13 to 16, as an in vitro diagnostic tool for the detection, by hybridization experiments, of nucleic acid sequences coding for a polypeptide according to any one of Claims 1 to 4, in biological samples, or for
- 10 the demonstration of aberrant syntheses or of genetic abnormalities.
23. Method of in vitro diagnosis for the detection of aberrant syntheses or of genetic abnormalities in the nucleic acid sequences coding for a polypeptide according to any one of Claims 1 to 4, characterized in that it comprises:
- 15 - the bringing of a nucleotide probe according to any one of Claims 13 to 16 into contact with a biological sample under conditions permitting
- 20 the formation of a hybridization complex between the said probe and the abovementioned nucleotide sequence, where appropriate after a prior step of amplification of the abovementioned nucleotide sequence;
- 25 - the detection of the hybridization complex possibly formed;
- where appropriate, the sequencing of the nucleotide sequence forming the hybridization complex with the probe of the invention.
- 30 24. Use of a nucleic acid sequence according to any one of Claims 6 to 8, for the production of a recombinant polypeptide according to any one of Claims 1 to 5.
25. Method of production of a recombinant SR-p70
- 35 protein, characterized in that transfected cells according to Claim 10 or 11 are cultured under conditions permitting the expression of a recombinant polypeptide of sequence SEQ ID No. 2, SEQ ID No. 4, SEQ ID No. 6, SEQ ID No. 8, SEQ ID No. 10, SEQ ID No. 13,
- 40 SEQ ID No. 15, SEQ ID No. 17 or SEQ ID No. 19 or any

biologically active fragment or derivative, and in that the said recombinant polypeptide is recovered.

26. Mono- or polyclonal antibodies or their fragments, chimeric antibodies or immunoconjugates, characterized in that they are capable of specifically recognizing a polypeptide according to any one of Claims 1 to 4.

27. Use of the antibodies according to the preceding claim, for the purification or detection of a polypeptide according to any one of Claims 1 to 4 in a biological sample.

28. Method of in vitro diagnosis of pathologies correlated with an expression or an abnormal accumulation of SR-p70 proteins, in particular the phenomena of carcinogenesis, from a biological sample, characterized in that at least one antibody according to Claim 25 is brought into contact with the said biological sample under conditions permitting the possible formation of specific immunological complexes between an SR-p70 protein and the said antibody or antibodies, and in that the specific immunological complexes possibly formed are detected.

29. Kit for the in vitro diagnosis of an expression or an abnormal accumulation of SR-p70 proteins in a biological sample and/or for measuring the level of expression of these proteins in the said sample, comprising:

- at least one antibody according to Claim 25, optionally bound to a support,
- means of visualization of the formation of specific antigen-antibody complexes between an SR-p70 protein and the said antibody, and/or means of quantification of these complexes.

30. Method for the early diagnosis of tumour formation, characterized in that autoantibodies directed against an SR-p70 protein are demonstrated in a serum sample drawn from an individual, according to the steps that consist in bringing a serum sample drawn from an individual into contact with a polypeptide of the invention, optionally bound to a support, under

conditions permitting the formation of specific immunological complexes between the said polypeptide and the autoantibodies possibly present in the serum sample, and in that the specific immunological complexes possibly formed are detected.

31. Method of determination of an allelic variability, a mutation, a deletion, an insertion, a loss of heterozygosity or a genetic abnormality of the SR-p70 gene, characterized in that it utilizes at least one nucleotide sequence according to any one of Claims 6 to 8.

32. Method of determination of an allelic variability of the SR-p70 gene at position -30 and -20 relative to the initiation ATG of exon 2 which may be involved in pathologies, and characterized in that it comprises at least:

- a step during which exon 2 of the SR-p70 gene carrying the target sequence is amplified by PCR using a pair of oligonucleotide primers according to any one of Claims 6 to 8;
- a step during which the amplified products are treated with a restriction enzyme whose cleavage site corresponds to the allele sought;
- a step during which at least one of the products of the enzyme reaction is detected or assayed.

33. Pharmaceutical composition comprising as active principle a polypeptide according to any one of Claims 1 to 4.

34. Pharmaceutical composition according to the preceding claim, characterized in that it comprises a polypeptide according to Claim 2.

35. Pharmaceutical composition containing an inhibitor or an activator of SR-p70 activity.

36. Pharmaceutical composition containing a polypeptide derived from a polypeptide according to any one of Claims 1 to 5, characterized in that it is an inhibitor or an activator of SR-p70.

1 TGCCTCCCGCCCCGCGCACCGCCCCGAGGCTGTGCTCCTGCGAAGGGG 50  
 1 .....GGGGCTCCGGGG 12  
 51 ACGCAGCGAAGCCGGGGCCCGGCCAGGCCGCGCGGACCGACGCCGATG 100  
 13 ACACCTGGCGTCCGGGCTGGAAGCGTGCTTCCAAGACGGTGACACGCTT 62  
 101 CCGGAGCTGCGACGGCTGCAGAGCGAGCTGCCCTCGGAGGCCGGTGTGA 150  
 63 CCGTGAGGATTGGCAGCCAGACTGCTTACGGGTAC...TGCCATGGAGG 109  
 151 GGAAGATGGCCAGTCCACCACCACTCCCGGATGGGGGACCAAGTTT 200  
 110 AGCCGCAGTCAGATCCAGCATCGAGCCCCCTCTGAGTCAGGAACAATT 159  
 201 GAGCACCTCTGGAGCTCTCTGGAAACGACAGCACTACTTCGACCTTC 250  
 160 TCAGACCTATGGAAACTACTTCTGAAACAAC.GTTCTGTCGCCCTTGC 208  
 251 CCAGTCAAGCCGGGGGAATATGAGGTGGTGGGTGGCAGGATTCACGA 300  
 209 CGTCCCAAGCGGTGGATGATTTGATGCTCTCTCCGGATGATTTGACAA 258  
 301 TGGACGTCTTCCACCTAGAGGGCATGACCACATCTGTGTCGCGCAGTTC 350  
 259 TGG.....TTAACTGAAGACCCAGGTC 280  
 351 AATTGCTGAGCAGCACCATGGACAGATGAGCAGCCCGCTGCCCTCGGC 400  
 281 CAGATGAAGCTC.....CCAGAAATGTGAGAGGCTGTCTCCCA 319  
 401 CAGCCCGTACACCCCGGAGCAGCCCGCCAGCGTCCCAACCAATTCACCT 450  
 320 TGGCCCCACACACCAGCAGCTCTACACCGGCGGCCCTGCACCAAGCCCC 368  
 451 ACGCACAGCCAGCTCCACCTTCGACACCATGTGCGCCCGGCGCTGTCTC 500  
 369 .....CTCCTGGCCCTGTCTCTCTGTC 393  
 501 CCGTCCAAACACGACTATCCCGGACCCCACTTCGAGGTCACTTTCCA 550  
 394 CCTTCCAGAAAACCTACCACGGCAGCTACGGTTTCCGCTGGGCTTCCT 443  
 551 GCAGTCCAGCAGGCCAAGTCAGCCACCTGGACGTACTCTCCCACTCTTGA 600  
 444 GCATTCTGGAAACAGCCAAAGTCTGTGACTTGCAGGTACTCCCTGACCTCA 493  
 601 AGAAAATCTACTGCGCAGATCGCCAAAGATGCCCCATCCAGATCAAGGTG 650  
 494 ACAAGATGTTTTGCGAGCTGGCCAAGACCTGCCCGTGCAGCTGTGGTT 543  
 651 TCCGCCCCACCGCCCCCGGGCACCGCCATCCGGGCCATGCGTGTACAA 700  
 544 GATTCCACACCCCGCGCGGCGAGCGCGCTCGCGCCATGGCCATCTACAA 593  
 701 GAAGCGGAGCAGCTGACCGACATCGTGAAGCGCTGCCCAACCAAGAGC 750  
 594 GCAGTACAGCACATGACTGAGGTGTTGAGGCGCTGCCCAACCAAGAGC 643  
 751 TCGGGAGGGACTTCAACGAAGGACAGTCTGCCCAAGCAGCCACCTCATC 800  
 644 GCTGCTCAGACAGCGATGGA.....CTGGCCCTCTCAACATCTTATC 687  
 801 COTGTGGAAGGCAATATCTCTCGCAGTATGTGACGACCCCTGTACCGG 850  
 688 CGAGTGAAGGAAATTTGGGTGTGAGTATTTCGATGACAGAAACACTTT 737  
 851 CAGGCAGAGCGTCGTGGTGCCCTATGAGCCACCAAGGTGGGACAGAAAT 900  
 738 TCGACATAGTGTGGTGGTCCCTATGAGCGGCTGAGGTGGCTGTGACT 787

FIG.1

FIG.1 cont.

```

1 MAQSTTTSPDGGTTFEHLWSSLEPDSTYFDLPQSSRGNNVVGSTSSMD 50
1 .....MEEPQSDPSIEPPLS....QETPSDLWKLLPENNVLSPLPQAVD 41
51 VFHLEGMTTSSVMAQFNLLSSTMDQMSRAASAPYTPPEHAASVPTHSPYA 100
42 DLML...SPDDLAQWLTEDPGPDEAPRMSEAAPHMAPTFAAPTFA..APAP 87
101 QPSSTFDTMSAPVIPSNTDYPGPHHFVFTFQSSSTAKSATWTYSPLLKK 150
88 APSWPL.....SSSVPSQKTYHGSYGFRLGFLHSGTAKSVTCTYSPDLNK 132
151 LYCQIAKTCPIQIKVSAPPPPGTAIRAMPVYKKAHVTDIVKRCPNHEL 200
133 MFCQLAKTCPVQLWVDSTPPPGSRVRMAIYKQSQHMTFVVRCPHHE.. 180
201 RDNFEGQSAPASHLIRVEGNLSQYVDDPVTGRQSVVVPYEPPOVGTEFT 250
181 RCDSDSGLAPPQHLIRVEGNLRVEYSDDRNTFRHSVVVPYEPPEVGSCT 230
251 TILYNFMCNSSCVGGMNRRPILIIITLETDRDQVLGRRSFEGRICACPR 300
231 TIHYNFMCNSSCMGMNRRPILTIITLEDSSGNLLGRNSFEVRVCACPR 280
301 DRKADEHYREQALNESSAKNGAASKRAFQSPPAVPALGPGVKRRHG 350
281 DRRTEENFRKKG...EPCHELPPGSTKRALPNNTSSSPQ.....PKKKPL 323
351 DEDTYLQVRGRENFEILMKLKESELMELVPQPLVDSYRQQQQLLRPS 400
324 DGEYFTLQIRGEREFEMFRELNEALELDAQAGKEPAGSRAHSSHLKSKK 373
401 HLQPPSYGPVLSPMNKVHGGVNKLPSVNQLVGQPPHSSAATPNLGPVGS 450
374 GQSTSRHKKFMFKTEGPDSD..... 393

```

FIG. 2

```

1  TGCTCTCCCGCCCGCGGCACCCGCGCCCGAGGCGCTGTGCTCTGCGAAGGGG 50
1  TGCTCTCCCGCCCGCGGCACCCGCGCCCGAGGCGCTGTGCTCTGCGAAGGGG 50
51  ACGCAGCGAAGCCCGGGGCCCGCGCCAGGCCGGGACGGACGCCGATG 100
51  ACGCAGCGAAGCCCGGGGCCCGCGCCAGGCCGGGACGGACGCCGATG 100
101  CCGGAGCTGCGACGCGCTGCAGAGCGAGCTGCCCTCGGAGCGCGGTGGA 150
101  CCGGAGCTGCGACGCGCTGCAGAGCGAGCTGCCCTCGGAGCGCGGTGGA 150
151  GGAAGATGGCCAGTCCACACCACTCTCCCGATGGGGGACCAACGTTT 200
151  GGAAGATGGCCAGTCCACACCACTCTCCCGATGGGGGACCAACGTTT 200
201  GAGCACCCTCTGGAGCTCTCTGGAAACGACAGCAGCACTACTTCGACCTTCC 250
201  GAGCACCCTCTGGAGCTCTCTGGAAACGACAGCAGCACTACTTCGACCTTCC 250
251  CCAGTCAAGCCGGGGGAATAATGAGGTGGTGGGTGGCACGGATTCCAGCA 300
251  CCAGTCAAGCCGGGGGAATAATGAGGTGGTGGGTGGCACGGATTCCAGCA 300
301  TGGACGCTCTTCCACCTAGAGGGCATGACCACATCTGTATCGGCCAGTTT 350
301  TGGACGCTCTTCCACCTAGAGGGCATGACCACATCTGTATCGGCCAGTTT 350
351  AATTTCGTGAGCAGCAGCATGGACGAGATGAGCAGCGCGCTGCCTCGGC 400
351  AATTTCGTGAGCAGCAGCATGGACGAGATGAGCAGCGCGCTGCCTCGGC 400
401  CAGCCCGTACACCCCGGAGCAGCGCCGAGCGTGCCACCCATTCACCTT 450
401  CAGCCCGTACACCCCGGAGCAGCGCCGAGCGTGCCACCCATTCACCTT 450
451  ACGCAGCGCCAGCTCCACCTTCGACACCATGTGCGCCGCGCTGTCAATC 500
451  ACGCAGCGCCAGCTCCACCTTCGACACCATGTGCGCCGCGCTGTCAATC 500
501  CCTTCCAACACCGACTATCCCGGACCCACCACTTCGAGGTCACTTTCCA 550
501  CCTTCCAACACCGACTATCCCGGACCCACCACTTCGAGGTCACTTTCCA 550
551  GCAGTCCAGCAGCGCCAAATCAGCCACCTGGACGTACTCCCCACTCTTGA 600
551  GCAGTCCAGCAGCGCCAAATCAGCCACCTGGACGTACTCCCCACTCTTGA 600
601  AGAAACTCTACTGCGCAGATCGCCAAGACATGCCCATCCAGATCAAGGTG 650
601  AGAAACTCTACTGCGCAGATCGCCAAGACATGCCCATCCAGATCAAGGTG 650
651  TCAGCCCAACCGCCCCCGGGCACCGCATCCGGGCGATGCTGTCTACAA 700
651  TCAGCCCAACCGCCCCCGGGCACCGCATCCGGGCGATGCTGTCTACAA 700
701  GAAAGCGGAGCAGCTGACCGACATCGTGAAGCGCTGCCCAACACGAGGC 750
701  GAAAGCGGAGCAGCTGACCGACATCGTGAAGCGCTGCCCAACACGAGGC 750
751  TCGGAGGGGACTTCAACGAAGGACAGTCTGCCCCAGCCAGCCACTCATC 800
751  TCGGAGGGGACTTCAACGAAGGACAGTCTGCCCCAGCCAGCCACTCATC 800
801  CGTGTGGAAGGCAATAATCTCTCGCAGATATGTGGACGACCTGTCAACGG 850
801  CGTGTGGAAGGCAATAATCTCTCGCAGATATGTGGACGACCTGTCAACGG 850
851  CAGGCAGAGCGTCTGGTGCCTATGAGCCACCAAGGTGGGACAGAAAT 900
851  CAGGCAGAGCGTCTGGTGCCTATGAGCCACCAAGGTGGGACAGAAAT 900

```

---

FIG.3  
cont.

FIG.3  
cont.

6/36

```

-- 1657 GATCCCCGAGCAGTATCGCATGACCATCTGGCGGGGCTGCAGGACCTGA 1706
-- |||
-- 1801 AGCAGGGCCACGACTACGGCGCCGCGCGCAGCAGCTGCTCCGCTCCAGC 1850
-- |||
-- 1707 AGCAGGGCCACGACTACGGCGCCGCGCGCAGCAGCTGCTCCGCTCCAGC 1756
-- |||
-- 1851 AACGCGGGCGCCATTTCATCGGCGGCTCCGCGGAGCTGCAGCGCCAGCG 1900
-- |||
-- 1757 AACGCGGGCGCCATTTCATCGGCGGCTCCGCGGAGCTGCAGCGCCAGCG 1806
-- |||
-- 1901 GGT CATGGAGGCCGTGC ACTTCCGCGTGCGCCAGACCATCACCATCCCCA 1950
-- |||
-- 1807 GGT CATGGAGGCCGTGC ACTTCCGCGTGCGCCAGACCATCACCATCCCCA 1856
-- |||
-- 1951 ACCGCGGCGGCCCCGGCGCCGGCCCCGACGAGTGGGCGACTTCGGCTTC 2000
-- |||
-- 1857 ACCGCGGCGGCCCCGGCGCCGGCCCCGACGAGTGGGCGACTTCGGCTTC 1906
-- |||
-- 2001 GACCTGCCCCGACTGCAAGGCCCGCAAGCAGCCCATCAAGGAGGAGTTCAC 2050
-- |||
-- 1907 GACCTGCCCCGACTGCAAGGCCCGCAAGCAGCCCATCAAGGAGGAGTTCAC 1956
-- |||
-- 2051 GGAGGCCGAGATCCACTGAGGGGGCGGGGCCAGCCAGAGCCTGTGCCACC 2100
-- |||
-- 1957 GGAGGCCGAGATCCACTGAGGGGGCGGGGCCAGCCAGAGCCTGTGCCACC 2006
-- |||
-- 2101 GCCCAGAGACCCAGGCCGCTCGCTCTC 2128
-- |||
-- 2007 GCCCAGAGACCCAGGCCGCTCGCTCTC 2034

```

FIG.3 cont.

7/36

00125005-072000

```

1  TGCCTCCCGCCCGCCGACCCGCGCCGAGGCGCTGTGCTCTCGAAGGGGACGACGGAA 60
61  GCCCGGGCCCGGCCAGGCGCGCGCGGACGGACGCGATGCCCGGAGCTGCCAGCGCTGC 120
121  AGAGCGAGCTGCCCTCGGAGCGCGGTGTGAGGAAGATGCCCGAGTCACCAACCACTCC 180
-10  M A Q S T T T S P 9
181  CCGATGGGGGACACGTTTGAACACCTCTGGAGCTCTCTTGAACACGACAGCACTACT 240
10  D G G T T F E H L M S S L E P D S T Y F 29
241  TCGACCTTCCCGAGTCAAGCCGGGGGAATAAGTGGTGGTGGTGGCAGCGATTCCAGCA 300
30  D L P Q S S R G N N E V V G G T D S S M 49
301  TGAAGCTCTTCCACCTAGAGGGGATGACACATCTGTCTATGCCCGAGTCAATTGTGTA 360
50  D V F H L E G H T T S V M A Q F N L L S 69
361  CGAGCACCATGGACAGATGAGCAGCGCGCGCTGCTCGGCCACCGCTACACCCCGGAGC 420
70  S T M D Q M S S R A A S A S P Y T P E H 69
421  ACGCCGCCAGCGCTGCCCAACCATTCACCTACGACAGCCGAGCTCCACCTTCGACACCA 480
90  A A S V P T H S P Y A Q P S S T F D T M 109
481  TGTCCGCGCGCTGTCTATCCCTTCAACCGACTATCCCGAGCCCAACCACTCGAGG 540
110  S P A P V I P S N T D Y P G P H H T E V 129
541  TCACTTTCAGCAGTCCAGCAGCGGCAAGTACGACCTGGACGTACTCCCACTCTGA 189
130  T F Q Q S S T A K S A T W T Y S P L L K 149
601  AGAACTCTACTGCCAGATCGCCAGACATGCCCATCCAGATCAAGGTGTCCGCGCCAC 660
150  K L Y C Q I A K T C P I Q I K V S A P P 169
661  CGCGCCCGGGACCGGCATCCGGCCATGCTGTCTACAAAGCGCGAGCACTGACCG 720
170  P P G T A I R A N P V V K K A E H V T O 189
721  ACATGCTGAAGCGCTGCCCAACACGAGCTCGGGAGGACTTCACGAAGACAGCTCTG 780
190  I V K R C P N H E L G R D F N E G Q S A 209
781  CCCGAGCAGCACCTCATCTCGTGGAGGCAATATCTCTCGCAGATGTGGAGCAGCC 840
210  F A S H L E N E G M N H L S Q I V D D P 229
841  CTGTCAACCGGACGAGAGGCTGTGTGTCCCTATGAGCCACACAGGTGGGGACAGAA 900
230  V T G R Q S V V V V P Y E P P Q V G T E F 249
901  TCACCACTCTGTCTACAACCTCATGTGTAAACGAGCTGTGTGGGGGCAATGAACGAC 960
250  T T I L Y N F M C N S S C V G G M N R R 269
961  GGCCCATCTCATCATCATCACCCTGGAGACCGGGATGGGAGGCTGTGGGCGCGCGGT 1020
270  P I L I T I T L E T R D G Q V L G R S 289
1021  CTTTGGAGGCGCATCTCGGCTGTCTCGCCCGCAGCGAAAGCGGATGAGGACCACT 1080
290  F E G R I C A C P G R D R K A D E D H Y 309
1081  ACCGGGAGCAGCGGCTTGAATGAGAGCTCGCGCAAGAACGGGCTCGCCAGCAGCGCG 1140
310  R E Q I A G N E G A N G C A A S K R A 329
1141  CTTTCAAGCAGATCCCCCTGCGCTCGCCCGCTGGGGCGCGGTTGAAAGACGAGCGGC 1200
330  F K Q S P P A V P A L G P G V K K R R H 349
1201  ACGGAGCAGGAGACACCTACTACTCTGAGGTGGAGGCGCGGAGAACTTCGAGATCTGA 1260
350  G E D E T Y Y L Q V R G R E N F E I L M 369
1261  TGAAGCTGAAGGAGGCTGAGAGCTGATGAGGTGTGTGCGCGAGCGCTGTGAGACTGCT 1320
370  K L K E S L E L M F L V Q F L V D S Y 389
1321  ATCGGACGAGCAGCAGCTCTACAGAGGCGGAGTCACTACAGCCCACTCTCTACGCGC 1380
390  R Q Q Q Q L L Q R P S H L Q P P S Y G 409
1381  CGGTCTCTCGCCCATGAAAGGTGACGCGGGGCGGTGAACAGCTGCGCTCGTCAACC 1440
410  V L S P M N K V H G G V N K L P S V N Q 429
1441  AGCTGTGGCGGAGCTCGCCCGCAGCTCTGAGCTACACCAACCTGGAGCACTGTG 1500
430  L V G Q P P P P H S S A A T P N L G P V G 449
1501  GCTCTGGATGCTCAACCAACCGGCGCAGCAGTGCACGCAACAGCGAGATGACAGCA 1560
450  S G M L N N H G H A V P A N S E M T S S 469
1561  GCGACGCGACCACTGCTATGGTCTCGGGGTCACCTGCACTCCGACCCGCTTACCACT 1620
470  H G T Q S H V S G S H C T P P P P Y H A 489
1621  CCGACGCCAGCTCTGCTAGTTTTTAAAGGATTGGGGTGTCAAAGCTGCTAGGATT 1680
490  D P S L V S F L T G L G C P N C I E Y P 509

```

---

FIG.4

```

1681 TCACGTCCAGGGGTTACAGAGCATTTACCACTGCGAGAACCTGACCATCGAGGACCTGG 1740
1510 T S Q G L Q S I Y H L Q N L T I E D L G 1729
1741 GGGCCCTGAAGATCCCCGAGCAGTATCGCATGACCATCTGGCGGGGCTGCGAGGACCTGA 1800
530 A L K I P E Q Y R M T I W R G L Q D L K 549
1801 AGCAGGGCCACGACTACGGCGCGCGCGCGCAGCAGCTGCTCCGCTCCAGCAACGCGCGCG 1860
559 Q G H D Y G A A A Q Q L L R S S H A A A 569
1861 CCATTTCATCGGGGCTCCGGGGAGCTCGAGCGCCAGCGGGTCATGGAGGCCCTGCACCT 1920
570 I S I G G S G E L Q R Q R V M E A V H F 589
1921 TCCGGCTGCGGCCACACCATCACCATCCCAACCGCGGGCGCGCGCGCGCGCGCGCGCG 1980
590 R V R H T I T I P N R G G P G A G P D E 609
1981 AGTGGGCGGACTTCGGCTTCGACCTGCGCGGACTGCAAGGCGCGCAAGCAGCCATCAAGG 2040
610 W A D F G F D L P D C K A R K Q P I K E 629
2041 AGGAGTTCACGGAGGCGGAGATCCCACTGAGGGGCGGGGCCAGCCAGAGGCTGTGCCACC 2100
630 E F T E A E I H * 649
2101 GCCCAGAGACCCAGGCGCGCTCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2160
1515 GGCCCTCCAGGCTGTGCGCGGGGAAAGGCAAGTCCCGGCCATGCGCCCGCACCTCACCGG 2220
2221 CCCACGAGAGGGCCGAGCCACCAAGCGCCCTCGCGGACAGCTGTAGTCACTGCAGAAC 2280
2281 TTCGGAGCTGCCCTAATGCTGGGCTTGGCGGGGAGGGGCGCGGCCACTCTCAGCCCTGC 2340
2341 CACTGCGGGGCTGCTCCATGGCAGCGCTGGGTGGGGACCGCAGTGTCTCAGCTCCGACCTC 2400
2401 CAGGCCCTCATCTAGAGACTCTGTCACTCTGCCGATCAAGCAAGTCTCTCCAGAGGAAAG 2460
2461 AATCTCTCTCTCTGTGGACTGCGCAAAAAGTATTTTGGGACATCTTTTGGTCTCTGGAGAG 2520
2521 TGGTGAGCAGCCAGGCGACTGTGTCTGAACACCTCTGATTTTCAAGGATGTCTCTTAC 2580
2581 GGGCTGGGGACTCTCTCTGTGGACTTGGAGTGTGCTTTTGGCCCGAGCACACTGTATTTC 2640
2641 TGGCGGAGCCGCTCTCTCTCTGCCCCCTAAACAACCAACAAAGTGTGTCTGAAATTGGAGAAA 2700
2701 ACTGGGGAAAGGCGCAACCCCTCCAGGTGCGGGGAAGCATCTGGTACCGGCTCGGCGAGT 2760
2761 CCCTCTAGGCTGGGCAAGTCACTCTCTCTCTGGGGAAACCTGGGCGAGAAAGGACAGCCT 2820
2821 GTCCTTAGAGGACCGGAAATTGTCAATATTGTATAAAATGATACCCCTTTCTAC 2874

```

FIG.4 cont.

1 TGCCCTCCCCCGCCGACCCGCGCCGAGGCGTGTGCTCTCGGAAGGGGACGCAGCGAA 60  
 61 GCGCGGGGCGCGCGCAGGCGCGCGCGGACCGACCGCGATGCGCGAGCTGCGACGGCTGC 120  
 121 AGAGCGAGTGCCTCCCTCGGAGCGCGGTGTGAGGAGATGGCCCACTCCACACCCACTCCCC 180  
 -10 CCGATAGCGGGGACACAGCTTTGAGCAGCTCTGGAGCAGACAGCAGCACTTACT 9  
 181 D G G T T F E H L W S S L E P D S T 29  
 10 D G G T T F E H L W S S L E P D S T 29  
 241 TCGACCTTCCCCAGTCAAGCGCGGGGAAATAGAGGTGTGGTGGCAGCGATTCCAGCA 300  
 30 D L P Q S S R G N N E V V G G T D S S M 49  
 301 TGGACGTCTTCCACCTAGAGGGCATGACCACATCTGTGATGGCCAGTTCAATTTCGTGA 360  
 50 D V F H L E G M T T S V N A Q F N L L S 69  
 361 GCAGCACCATGAGCCAGATGAGCAGCGCGCTGCTCGGCCAGCCGTACACCCCGGAGC 420  
 70 S T M D Q M S S R A A S A S P Y T P E H 89  
 421 AGCGCGCCGAGCGTGCACCCATTCACCTTACGACAGCCAGCTCCACCTTCGACACCA 480  
 90 A A S V P T H S P Y A Q P S S T F D T M 109  
 481 TGTGGCCGCGCGCTGTCTATCCCTCGAAGCAGCACTATCCCGGACCCACACCTTCAGG 540  
 110 S P A P V I P S N T D Y P G P H H F E V 129  
 541 TCACCTTCCAGCAGTCCAGCAGCGCCAGTCCAGCCACTGGACGTACTCCCACTCTTGA 600  
 130 T F Q Q S S T A K S A T W T Y S P L L K 149  
 601 AGAAACTCTACTGCCAGATCGCCAGACATGCGCCATCGAGTACAGGTGTGCGGCCAC 660  
 150 K L Y C Q I A K T C P I Q I K V S A P P 169  
 661 CCGCCCCGGGACCGCCATCCGGGCGATGCTCTTACAGAAAGCGGAGCAGCTGACCG 720  
 170 P P G T A I R A M P V Y K K A E H V T D 189  
 721 ACATCGTGAAGCGCTGCCCAACCAAGCAGCTCGGGAGGCACTTCAAGGAGGACAGTCTG 780  
 190 I V K R C P N H E L G R D F N E G Q S A 209  
 781 CCCCAGCCAGCCACCTCATCGGTGTGGAAGGCAATATCTCTCCAGTATGTGGAGCACC 840  
 210 P A S H L I R V E G N N L S Q Y V D D P 229  
 841 CTGTACCGCGGAGGACAGCGCTGTGTGCGCTATGAGCCACCAAGGTGGGGACAGAA 900  
 230 V T G R Q S V V P P Y E P P Q V G T E F 249  
 901 TCACCAACCATCTGTACAACTTCAATGTGTAAACAGCAGCTGTGTGGGGGGCATGAACCG 960  
 250 T T I L Y N F M C N S S C V G G M N R R 269  
 961 GGCCATCTCTATCATCATCACTCCAGAGCAGGATGGCGAGTGTCTGGGCGCCGGT 1020  
 270 P I L I I I T L E T R D G Q V L G R R S 289  
 1021 CTTTGGAGGCGGCACTCTGCGCCTGTCTCTGCGCGGACCGAAAGCCGATGAGGACACT 1080  
 290 F E G R I C A C P G R D R K A D E D H Y 1140  
 1081 ACCGGGAGCAGCAGCGCTTGAATGAGAGCTCCGCCAAGAACCGGGCTGCCAGCAAGCGGG 1140  
 310 R E Q Q A L N E S S K N G A A S K R A 329  
 1141 CTTTCAAGCAGAGTCCCGCTGCGCTCCCGCTGCGCCCGGTGTGAAGAAGCGCGCGC 1200  
 330 F K Q S P P A V P A L G P G V K K R R H 349  
 1201 ACGGAGCAGGAGCAGCTACTACTCTGAGGTGCGAGGCGCGGAGAATTTCGAGATTCCTGA 1260  
 350 G D E D T Y Y L Q V R G R E N F E I L M 369  
 1261 TGAAGCTGAAGGAGAGCTCTGAGCTGATGAGATGTGTCTCCAGCCGCTGTGAGACTCTC 1320  
 370 K L K F S L E L M E L V P Q P L V D S Y 389  
 1321 ATCGGAGCAGCAGCAGCTCTCTACAGAGGCGAGTCACTACAGCCCCCACTCTCAGGGCG 1380  
 390 R Q Q Q Q L L L Q R P S H L Q P P S Y G P 409  
 1381 CGGTCTCTCTGCCCCATGAACAAGGTGACCGGGGGGTGAACAAGCTGCCTCCGTCAACC 1440  
 410 V L S P M N K V H G G V N K L P S V N Q 429  
 1441 AGCTGTGTGGCCAGCCTCTCCCGCAGCAGCTGGCACTACACCAAGCTGGAGCTGTGG 1500  
 430 L V G Q P P P H S S A A T P N L G P V G 449  
 1501 GCTCTGGGATGCTCAACACCAACCGCCACCGAGTGGCGGCCAACAGCGAGATGACACGA 1560  
 450 S G M L N N H G H A V P A N S E M T S 469  
 1561 GCGACGGCACCCAGTCCATGGTCTCGGGTTCGACGTACTCCGACCCCTTACACAG 1620  
 470 H G T Q S H V S G S H C T P P P P Y H 489  
 1621 CCGACCCAGCCTCTCTGAGGACCTGGGGCGCTTGAAGTCCCGAGCAGTATCGCATGAC 1680  
 490 D P S L V R T W G P \* 509  
 1681 CATCGGGGGGGCTCGAGGACCTGAAGCAGGGGCCAGCTACCGCGCCGCGCGCGCAGCA 1740  
 1741 GCTCCTCGGCTCAGCAACCGCGCGCATTTCCATCGCGCGCTCGCGGAGCTGCAGAGC 1800  
 1801 CCGACGGTATCGAGGCGCTGCACTTCTCGGGCGGACAGCTTACAGTATCCCATCCCAACG 1860  
 1861 CGCGCGCCCGCGCGCGCGCCCGCAGTGTGGCGGCACTTGGCTTTCGACTCTGCGCGAC 1920  
 1921 CAAGCGGCGCAAGCAGCCCATCAAGGAGGAGTTTCAAGGAGCGGATCTCACTGAGGGG 1980  
 1981 CGGSGCCAGCAGAGGCTGTGCCACCGCCAGAGACCCAGGCGCGCTGTGCTCTC 2034

FIG.5

10/36

1 GCGAGCTGCCCTCGGAGCCGGCGTGGGGAAGATGGCCAGTCCACGCCACCTCCCCCTG 60  
 -9 MA Q S T A T S P D 10  
 61 ATGGGGGCACACGTTTGAACACCTCTGGAGCTCTCTGGAACAGACAGACCTACTCTG 120  
 11 G G T T F E H L W S S L E P D S T Y F D 30  
 121 ACCTTCCCCAGTCAAGCCGGGGGAATAATGAGGTGGTGGGCGGAACGGATTCCAGCATGG 180  
 31 L P Q S S R G N N E V V G G T D S S M D 50  
 181 ACGTCTTCCACCTGGAGGGCATGATACATCTGTCAATGAGTTCATCTGTGAGGA 240  
 51 V F H L E G M T T S V M A Q F N L L S S 70  
 241 GCACCATGAGCAGATGAGCAGCGCGCGGCTCGGCAGCCCTACACCCAGAGCAGC 300  
 71 T M D Q M S S R A A S A S P Y T P E H A 90  
 301 CGCCGAGCGTGGCCACCCACTCGCCCTACGCAACAACCCAGCTCCACCTTCGACACCATGT 360  
 91 A S V P T H S P Y A Q P S S T F D T M S 110  
 361 CGCCGGCGCCTGTATCCCTCCAAACACGACTACCCCGGACCCACCACCTTTGAGGTCA 420  
 111 P A P V I P S N T D Y P G P H H F E V T 130  
 421 CTTTCCAGCAGTCCAGCACGGCCAAAGTCAGCCACTGGACGTACTCCCGGCTCTTGAAGA 480  
 131 F Q Q S S T A K S A T W T Y S P L L K K 150  
 481 AACTCTACTGCCAGATCGCCAAGACATGCCCATTCAGATCAAGGTGTCCACCCCGCCAC 540  
 151 L Y C Q I A K T C P I Q I K V S T P P P 170  
 541 CCCCAGGCACTGCCATCCGGCCCATGCTGTATTACAAGAAAGCGAGCACGTGACCCGACG 600  
 171 P G T A I R A M P V Y K K A E H V T D V 190  
 601 TCGTGAACGCTGCCCAACACGAGCTCGGGAGGGACTTCAACGAAGGACAGTCTGCTC 660  
 191 V K R C P N H E L G R D F N E G Q S A P 210  
 661 CAGCCAGCCACTCATCCGCGTGAAGGCAATAATCTCTCGCAGTATGTGGATGACCCCTG 720  
 211 A S H L I R V E G N N L S Q Y V D D P V 230  
 721 TCACCGGCGGAGAGCGTCTGGTGGCCCTATGAGCCACCAAGTGGGGAGCGGAATTC A 780  
 231 T G R Q S V V V P Y E P P Q V G T E F T 250  
 781 CCACCATCTGTACAACTTCATGTGTAAACAGCAGCTGTGTAGGGGGCATGAACCGGGCG 840  
 251 T I L Y N F M C N S S C V G G M N R R P 270  
 841 CCATCTCATCATCATCACTCGGAGATCGGGATGGGCAGGTGCTGGGCGCGCGGTCT 900  
 271 I L I I I T L E M R D G Q V L G R R S C T 920  
 901 TTGAGGCGCCATCTGCGCCTGTCTGGCGCGGACCGAAAGCTGTAGGAGACCACTACC 960  
 291 E G R I C A C P G R D R K A D E D H Y R 310  
 961 GGGAGCAGCAGGCGCTGAACGAGAGCTCCGCCAAGAACGGGGCGCCAGCAAGCGTGCTC 1020  
 311 E Q Q A L N E S S A K N G A A S K R A F 330  
 1021 TCAAGCAGAGCCCCCTGCGCTGCCCGCCTTGGTGGCGGTGTGAAGAAGCGGGCGGCA 1080  
 331 K Q S P P A V P A L G A G V K K R R H G 350  
 1081 GAGAGGAGACAGCTACTACCTTCAAGTCCGAGGCGGGAGAACTTTGAGATCCTGTATGA 1140  
 351 D E D T Y Y L Q V R G R E N F E I L M K 370  
 1141 AGCTGAAAGAGAGCCTGGAGCTGATGGAGTTGGTGCCGAGCCACTGGTGAGACTCTATC 1200  
 371 L K E S L E L M E L V P Q P L V D S Y R 390  
 1201 GGCAGCAGCAGCTCCTACAGAGCGGAGTCACCTACAGCCCCGCTCCTACGGGCGGG 1260  
 391 Q Q Q Q L L Q R P S H L Q P P S Y G P V 410  
 1261 TCCTCTGCCCATGAACAAGGTGACCGGGGCATGAACAAGCTGCCCTCCGTCAACACAGC 1320  
 411 L S P M N K V H G G M N K L P S V N Q L 430  
 1321 TGTGGGCGCAGCTCCCCGCACAGTTCCGGCAGCTACACCAACCTGGGGCCGTGGGCC 1380  
 431 V G Q P P P H S S A A T P N L G P V P G 450  
 1381 CCGGAGTGCTCAACAAACCATGGCCACGAGTGCAGGCCAACGGCAGATGAGCAGCAGCC 1440  
 451 G M L N N H G H A V P A N G E M S S S H 470

FIG.6

1441 ACAGCGCCAGTCCATGGTCTCGGGGTCCCACTGCACCTCCGCCACCCCTACACGCCG 1500  
 471 S A Q S M V S G S H C T P P P P Y H A D 490  
 1501 ACCCCAGCCTCGTCAGTTTTTTAACAGGATTGGGGGTGCCAACTGCATCGAGTATTTCA 1560  
 491 P S L V S F L T G L G C P N C I E Y F T 510  
 1561 CCTCCCAAGGGTTACAGAGCATTACCACCTGCAGAACCTGACCATTGAGGACCTGGGGG 1620  
 511 S Q G L Q S I Y H L Q N L T I E D L G A 530  
 1621 CCCTGAAGATCCCCGAGCAGTACCGCATGACCATCTGGCGGGGCTGCAGGACCTGAAGC 1680  
 531 L K I P E Q Y R M T I W R G L Q D L K Q 550  
 1681 AGGGCCACGACTACAGCACCGCGCAGCAGCTGCTCCGCTCTAGCAACGCGGCCACCATCT 1740  
 551 G H D Y S T A Q Q L L R S S N A A T I S 570  
 1741 CCATCGGCGGCTCAGGGAACTGCAGCGCCAGCGGTCATGGAGGCCGTGCACCTCCCGG 1800  
 571 I G G S G E L Q R Q R V M E A V H F R V 590  
 1801 TGCGCCACACCATCACCATCCCCAACCGCGGCGGCCAGGCGCGGCCCTGACGAGTGGG 1860  
 591 R H T I T I P N R G G P G G G P D E W A 610  
 1861 CGGACTTCGGCTTCGACCTGCCGACTGCAAGGCCCGCAAGCAGCCCATCAAGGAGGAGT 1920  
 611 D F G F D L P D C K A R K Q P I K E E F 630  
 1921 TCACGGAGGCCGAGATCCACTGAGGGCCTCGCCTGGCTGCAGCCTGCGCCACCGCCAGA 1980  
 631 T E A E I H \* 650  
 1981 GACCCAAGTCGCCTCCCCTCCTTCCTGTGTGTCCAAAACCTGCCTCAGGAGGCAGGACC 2040  
 2041 TTCGGGCTGTGCCCCGGGAAAGGCAAGGTCCGGCCCATCCCAGGCACCTCACAGGCCCC 2100  
 2101 AGGAAAGGCCCAGCCACCGAAGCGCCTGTGGACAGCCTGAGTCACCTGCAGAAC 2156

FIG.6 cont.

1 TGATCTCCCTGTGGCCTGCAGGGGACTGAGCCAGGGAGTAGATGCCCTGAGACCCCAAGG 60  
 61 GACACCCAAGGAAACCTTGCTGGCTTGTGAGAAAGGATCGTCTCTCTGCCCAAGAGA 120  
 121 AGCATGTGTATGGGCGCCTGTGTATGAATCCTTGGGGCAGGCCCACTCAATTGCTCAGC 180  
 181 M C N G P V Y E S L G Q A Q F N L L S 19  
 191 ATGTCATGGACAGATGGCGAGCGTGGCGGCCCGGGGAGGCCCTACACCCCGAGGAC 240  
 20 S A M D Q M G S R A A P A T P E H 39  
 241 GCGCGCAGCGGCCCAACCCTCGCCCTACGGCAGCCCACTCCAGCTTCGACACCATG 300  
 40 A A S A P T H S P Y A Q P S S T F D T M 39  
 301 TCTCGGCGCCTGTGCATCCCTTCCAATACCGACTACCCCGGCCCCACACTTCGAGTGC 360  
 60 S P A P V I P S N T D Y P G P H H F E V 79  
 361 ACCTTCCAGCAGCTCGAGCAGTGCACAGTGGCCACCTGGACATCTCCCACTCTGGAAG 420  
 80 F C Q G S S T A K S A T W T Y S P L L K 99  
 421 AAGTTGTACTGTGAGATTGCTAAGACATGCCCATCTCCAGATCAAGGTGCCACCCACA 480  
 100 K L Y C Q T A K T C P T Q I K V S T 119  
 481 CCCCCGGGCAGGCCATCCGGGCCATGCTGTCTACAAAGAGGCAGACATGTGACCGAC 540  
 120 P P G T A I R A M P V V K K A E H V T D 139  
 541 ATTGTAAAGCGTGCCTCAACACGAGCTTGGAAAGGACTTCAATGAAGCAGCTCTGCC 600  
 140 I V K R C P N H E L G R D F N E G Q S 159  
 601 CCGGCTAGGCCACCTCATCCGTGTAGAAGGCAACAACCTGCCCCAGTACGTGGATGACCT 660  
 160 P A S H L I R V E G N N L A Q Y V D D P 179  
 661 GTGACCAAGGAGCAGAGTGTGGTGTGGCGTATGAACCCACAGGTGGGACGAATTT 720  
 180 V T G R S V P E P P V G G T E F 199  
 721 ACCACCATCTGTACAACTTCATGTGTAAAGCAGCTGTGGGGGCCATGAATCGGAGG 780  
 200 T T I L Y N F M C N S S C V G S M N 219  
 781 CCCATCTTGTTCATCATCACCCTGGAGACCCGGGATGGACAGGTCTGGCGCCCGCGTCT 840  
 220 P I L V I I T L E T R D G Q V L G R R S 239  
 841 TTCGAGGTTGCATCTGTGCTGTCTGTGGCGGTGACCGCAAGTGTGAAGACCAATTAC 900  
 240 F E G R I C A C P G R D R K A C D E D H Y 259  
 901 CGGAGCAACAGCGCTGTGAATGAAGTACCAACCAAAAATGGAGTGCACAGCAAGCGTGA 960  
 260 R E Q A L N E S T K N G A A S K R A 279  
 961 TTCACGAGAGCCCCCTGCCATCCCTGCCCTGGTACCAAGTGAAGAGAGCGCCAC 1020  
 280 F K Q S P P A I P A L G T N V K K R R H 1080  
 1021 GGGGACGAGGACATGTCTACATGCACGTGGGAGCGCGGAGAACTTGTGATCTTGATG 1140  
 300 G D E D M F Y M H V R G R E N F E I L M 319  
 1081 AAAGTC AAGGAGGCTAGAACTGATGGAGCTTGTGGCCGAGCCTTGGTTGACTCTGAT 1140  
 320 K V K E S L E L M E L V P Q P L V D S Y 339  
 1141 CGACAGCAGCAGCAGCAGCAGCTCTACAGAGGCGAGTCACTGCAGCCTCCATCTGAT 1200  
 340 R Q Q Q Q Q L L Q R P S H L Q P P S Y 359  
 1201 GGGCCCGTGTCTTCCCCAAATGAACAAGGTACACGCTGTGTCTCAACAACTGCCCTCCG 1260  
 360 G P V L S P M N K V H G V N K L F S 379  
 1261 AACCAGCTGGTGGGCGAGCTCCCGCCGACAGCTCAGCAGCTGGGCCCCAACCTGGGGCCC 1320  
 380 N Q L V G Q P P P H S S A A G P N L G P 399  
 1321 ATGGCTCCGGGATGCTCAACAGCCACGGCCACAGCATGCCGGCCAAATGGTGAGATGAAT 1380  
 400 M G S G M L N S H G H S M P A N G E M N 419  
 1381 GGAGGCCACAGCTTCCGACACCATGGTTTCGGGATCTCCACTGACCCCGCCACCCCTAT 1440  
 420 C G H S Q T H V S G S H C T P P P P Y 439  
 1441 CATGAGACCCAGCCTCTGTCAFTTTTGACAGGGTGGGGTGTCCAAACTGCATCGAG 1500  
 440 H A D P S L V S F L T G L G C P M C T E 459  
 1501 TGCTTCACTTCCCAAGGGTTGACAGAGCATCTACCACTGCAGAACTTACCATTGCAGGAC 479  
 460 C T T G G G C T C T G A A G G T C C T G A C C A G T A C C A T C T G G A G G G C T A C A G G A C 1620  
 1561 L G A L K V P D Q Y R M T I W R G L Q D 499  
 1621 CTGAAGCAGAGCCATGACTGGCGCCAGCAACTGTCTACGCTCCAGCAGCAACCGCGGCCACC 1680  
 500 L K Q S H D C G Q Q L L R S S N A A T 519  
 1681 ATCTCCATCGGGCTCTGGCGAGCTGCAGCGGCGAGGGTCTATGGAAGCGGTGCTATTTC 1740  
 520 I S T G G S F E L Q R G V M E A V H F 1800  
 1741 CGTGTGGCCCAACCATCACAATCCCCAACCTGTGAGGCGCAGGTGCGGTGACAGTCC 559  
 540 R V R H T I T I P N R G G A G A V T G P 579  
 1801 GACGAGTGGCGGACCTTGGCTTTCAGCTGCTGACTGCAAGTCCCGTAAGCAGCCATC 1860  
 560 D E W A D F G F D L P D C K S R K Q P I 579  
 1861 AAGAGGAGTTCACAGAGACAGAGGCCACTGAGAACCTACTCTCTCTCTGCTCTTC 1920  
 580 K E F T E S H 599  
 1921 CTCTGTGAGAACTGCTCTTGGAAAGTGGAGCTGTGTGGCTGTGCCACAGAAACAGCAA 1980  
 1981 GGACCTTCTCGGATGCCATTCCTGAAGGGAAAGTCCGTCATGACTAATCTCTCTTGG 2040

FIG.7

```

1  TGGTCCCGCTTCGACCAAGACTCCGGCTACCAAGCTTGC GGCGCCCGCGGAGGAGAGCC 60
61  CCGCTGGGGCTAGCTGGGCGACGCGCGCCAAGCGCGCGGGAAGGAGGCGGGAGGAGCG 120
121  GGGCCCGAGACCCCGACTCGGGCAGAGCCAGCTGGGAGGCGGGGCGCGCTGGGAGCCA 180
181  GGGGCCCGGGTGGCCGGCCCTCTCTCGCCACGGCTAGTCCCGCGCTGCTTCCCGCGG 240
241  GTCCGCCAAGAAAGGCGCTAAGCCTCGGGCAGTCCCTCGCGCGCGCTCCCTGCTCCGC 300
301  ACCCTTATAACCCGGCGTCCCGCATCGGCGGAGGAGCAACGCTGCAGCCAGCCCTCG 360
361  CCGACGCCGACGCCCGGCCGGAGCAGAATGAGCGCGAGCGTTGGGGAGATGGCCAGAC 420
-8  M S G S V G E M A Q T 11
421  CTCTTCTTCTCTCTCTCCACCTTCGAGCACCTGTGGAGTTCTCTAGAGCCAGACAGCAC 480
12  S S S S S S S T F E H L W S S L E P D S T 31
481  CTACTTTGACCTCCCCAGCCAGCCAGGGACTAGCGAGGCATCAGGCAGCGAGGAGTC 540
32  Y F D L P Q P S Q G T S E A S G S E E S 51
541  CAACATGGATGCTCTTCACCTGCAAGGCATGGCCCAAGTCAATTGCTCAGCAGTGGCAT 600
52  N M D V F H L Q G M A Q F N L L S S A M 71
601  GGACCAGATGGGCAGCGGTGCGGCCCGCGGAGCCCTACACCCGGAGCACGCGCCGAG 660
72  D Q M G S R A A P A S P Y T P E H A A S 91
661  CGCGCCCAACCCACTCGCCCTACGCGCAGCCAGCTCCACCTTCGACACCATGTCTCCGG 720
92  A P T H S P Y A Q P S S T F D T M S P A 111
721  GCCTGTATCCCTTCCAATACCGACTACCCCGGCCCC 758
112  P V I P S N T D Y P G P 123

```

FIG. 8

14/36

```

- Name: sr-p70a-cos3      Len:   650  Check: 9661  Weight: 1.00
- Name: sr-p70b-cos3      Len:   650  Check: 3605  Weight: 1.00
- Name: sr-p70-ht29       Len:   650  Check:   85  Weight: 1.00
- Name: sr-p70c-att20     Len:   650  Check: 4072  Weight: 1.00
- Name: sr-p70a-att20     Len:   650  Check: 4204  Weight: 1.00
-
-//
-
-      1
- sr-p70a-cos3      .....MAQ STTTSPDGGT TFEHLWSSLE PDSTYFDLPQ SSRGNNEVVG
- sr-p70b-cos3      .....MAQ STTTSPDGGT TFEHLWSSLE PDSTYFDLPQ SSRGNNEVVG
- sr-p70-ht29       .....MAQ STATTSPDGGT TFEHLWSSLE PDSTYFDLPQ SSRGNNEVVG
- sr-p70c-att20     .....TSSSSSS TFEHLWSSLE PDSTYFDLPQ PSQQTSEASG
- sr-p70a-att20     MSGSVGEMAQ .....TSSSSSS TFEHLWSSLE PDSTYFDLPQ PSQQTSEASG
-
-      51
- sr-p70a-cos3      GTDSSMD.VF HLEGMTTSMV AQFNLLSSTM DQMSRAASA SPYTPEHAAS
- sr-p70b-cos3      GTDSSMD.VF HLEGMTTSMV AQFNLLSSTM DQMSRAASA SPYTPEHAAS
- sr-p70-ht29       GTDSSMD.VF HLEGMTTSMV AQFNLLSSTM DQMSRAASA SPYTPEHAAS
- sr-p70c-att20     ...MCMGPVY ...ESLG...Q AQFNLLSSAM DQMSRAAPA SPYTPEHAAS
- sr-p70a-att20     SEESNMD.VF HLQGM..... AQFNLLSSAM DQMSRAAPA SPYTPEHAAS
-
-      101
- sr-p70a-cos3      VPTHSPYAQP SSTFDTMSPA PVIPSNTDYP GPHHFVETFO QSSTAKSATW
- sr-p70b-cos3      VPTHSPYAQP SSTFDTMSPA PVIPSNTDYP GPHHFVETFO QSSTAKSATW
- sr-p70-ht29       VPTHSPYAQP SSTFDTMSPA PVIPSNTDYP GPHHFVETFO QSSTAKSATW
- sr-p70c-att20     APTHSPYAQP SSTFDTMSPA PVIPSNTDYP GPHHFVETFO QSSTAKSATW
- sr-p70a-att20     APTHSPYAQP SSTFDTMSPA PVIPSNTDYP GP.....
-
-      151
- sr-p70a-cos3      TYSPLLKKLY CQIAKTCPIQ IKVSAPPPPG TAIRAMPVVK KAEHVTDIVK
- sr-p70b-cos3      TYSPLLKKLY CQIAKTCPIQ IKVSAPPPPG TAIRAMPVVK KAEHVTDIVK
- sr-p70-ht29       TYSPLLKKLY CQIAKTCPIQ IKVSTPPPPG TAIRAMPVVK KAEHVTDIVK
- sr-p70c-att20     TYSPLLKKLY CQIAKTCPIQ IKVSTPPPPG TAIRAMPVVK KAEHVTDIVK
- sr-p70a-att20     TYSPLLKKLY CQIAKTCPIQ .....
-
-      201
- sr-p70a-cos3      RCPNHLEGRD FNEGQSAPAS HLIRVEGNL SQYVDDPVFG RQSVVVPYEP
- sr-p70b-cos3      RCPNHLEGRD FNEGQSAPAS HLIRVEGNL SQYVDDPVFG RQSVVVPYEP
- sr-p70-ht29       RCPNHLEGRD FNEGQSAPAS HLIRVEGNL SQYVDDPVFG RQSVVVPYEP
- sr-p70c-att20     RCPNHLEGRD FNEGQSAPAS HLIRVEGNL SQYVDDPVFG RQSVVVPYEP
- sr-p70a-att20     .....
-
-      251
- sr-p70a-cos3      PQVGTEFTTI LYNFMCNSSC VGGMNRRLPIL IIITLETDRG QVLGRRSFEG
- sr-p70b-cos3      PQVGTEFTTI LYNFMCNSSC VGGMNRRLPIL IIITLETDRG QVLGRRSFEG
- sr-p70-ht29       PQVGTEFTTI LYNFMCNSSC VGGMNRRLPIL IIITLETDRG QVLGRRSFEG
- sr-p70c-att20     PQVGTEFTTI LYNFMCNSSC VGGMNRRLPIL IIITLETDRG QVLGRRSFEG
- sr-p70a-att20     .....
-
-      301
- sr-p70a-cos3      RICACPGDRR KADEHDHYREQ QALNESSAKN GAASKRAFPQ SPPAVPALGP
- sr-p70b-cos3      RICACPGDRR KADEHDHYREQ QALNESSAKN GAASKRAFPQ SPPAVPALGP
- sr-p70-ht29       RICACPGDRR KADEHDHYREQ QALNESSAKN GAASKRAFPQ SPPAVPALGA
- sr-p70c-att20     RICACPGDRR KADEHDHYREQ QALNESSAKN GAASKRAFPQ SPPAVPALGT
- sr-p70a-att20     .....

```

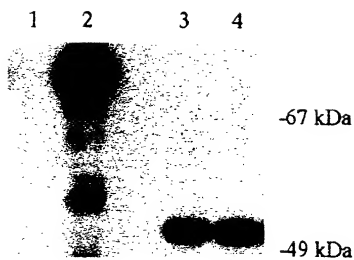
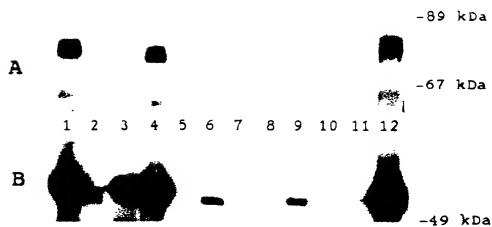
...

FIG. 9

351 400  
 - sr-p70a-cos3 GVKKRRHGD ETVYLVQRGR ENFEILMKLK ESLELMELVP QPLVDSYR..  
 - sr-p70b-cos3 GVKKRRHGD ETVYLVQRGR ENFEILMKLK ESLELMELVP QPLVDSYR..  
 - sr-p70-hc29 GVKKRRHGD ETVYLVQRGR ENFEILMKLK ESLELMELVP QPLVDSYR..  
 - sr-p70c-att20 NVKKRRHGD DMFYMHVRGR ENFEILMKLK ESLELMELVP QPLVDSYRQQ  
 - sr-p70a-att20 .....  
 -  
 401 450  
 - sr-p70a-cos3 QQQQLLQRP S HLQPPSYGPV LSPMNKVHGG VNKLP SVNQL VGQPPPHSSA  
 - sr-p70b-cos3 QQQQLLQRP S HLQPPSYGPV LSPMNKVHGG VNKLP SVNQL VGQPPPHSSA  
 - sr-p70-hc29 QQQQLLQRP S HLQPPSYGPV LSPMNKVHGG VNKLP SVNQL VGQPPPHSSA  
 - sr-p70c-att20 QQQQLLQRP S HLQPPSYGPV LSPMNKVHGG VNKLP SVNQL VGQPPPHSSA  
 - sr-p70a-att20 .....  
 -  
 451 500  
 - sr-p70a-cos3 ATPNLGPVGS GMLNNHGHAV PANSEMTSSH GTQSMVSGSH CTPPPPYHAD  
 - sr-p70b-cos3 ATPNLGPVGS GMLNNHGHAV PANSEMTSSH GTQSMVSGSH CTPPPPYHAD  
 - sr-p70-hc29 ATPNLGPVGS GMLNNHGHAV PANSEMTSSH GTQSMVSGSH CTPPPPYHAD  
 - sr-p70c-att20 AGPNLGPVGS GMLNNHGHAV PANSEMTSSH GTQSMVSGSH CTPPPPYHAD  
 - sr-p70a-att20 .....  
 -  
 501 550  
 - sr-p70a-cos3 PSLVSFLTGL GCPNCIEYFT SQGLQSIYHL QMLTIEDLGA LKIQEYRMT  
 - sr-p70b-cos3 PSLVR..T.W G.P.....  
 - sr-p70-hc29 PSLVSFLTGL GCPNCIEYFT SQGLQSIYHL QMLTIEDLGA LKIQEYRMT  
 - sr-p70c-att20 PSLVSFLTGL GCPNCIEYFT SQGLQSIYHL QMLTIEDLGA LKVPDQYRMT  
 - sr-p70a-att20 .....  
 -  
 551 600  
 - sr-p70a-cos3 IWRGLQDLKQ GHDYGAAQQ LLR.SSNA A ISIGSGSELQ RQRVMEAVHF  
 - sr-p70b-cos3 IWRGLQDLKQ GHDYS.TAQ LLR.SSNA A ISIGSGSELQ RQRVMEAVHF  
 - sr-p70-hc29 IWRGLQDLKQ GHDYS.TAQ LLR.SSNA A ISIGSGSELQ RQRVMEAVHF  
 - sr-p70c-att20 IWRGLQDLKQ SHDCG...QQ LLR.SSNA A ISIGSGSELQ RQRVMEAVHF  
 - sr-p70a-att20 .....  
 -  
 601 650  
 - sr-p70a-cos3 RVRHTITIPN RGGPGA..GP DEWADPGFDL PDCKARKQPI KEEFTEAETH  
 - sr-p70b-cos3 RVRHTITIPN RGGPGA..GP DEWADPGFDL PDCKARKQPI KEEFTEAETH  
 - sr-p70-hc29 RVRHTITIPN RGGPGA..GP DEWADPGFDL PDCKARKQPI KEEFTEAETH  
 - sr-p70c-att20 RVRHTITIPN RGGGAVTGP DEWADPGFDL PDCKARKQPI KEEFTEAETH  
 - sr-p70a-att20 .....  
 -

FIG.9 cont.

16/36

FIG.10aFIG.10b

17/36

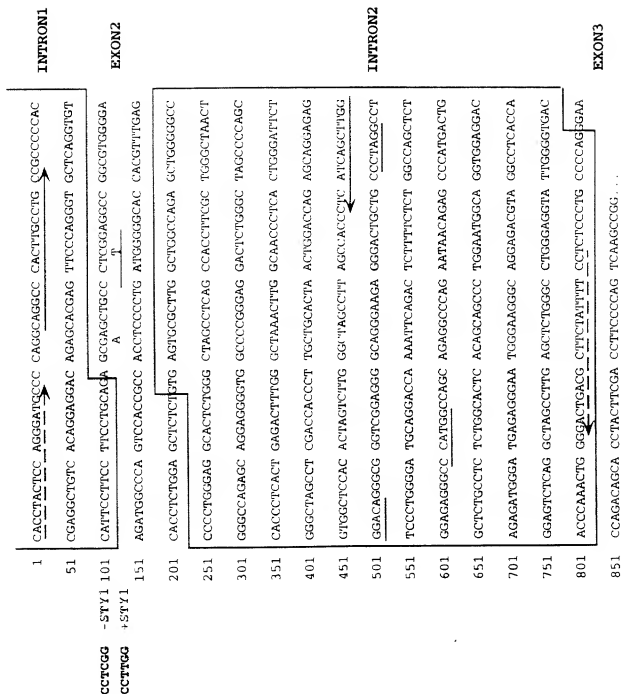


FIG.11

1 ↔ 2 1 MAQS..TATSPDGGTTFEHLWSSLEPDSTYFDLPQSSRGNNVEVGGTSSMD 50  
 1 ↔ 2 1 MEEPQSDPSVEPPLSQETTFSDLWKLLPE.....NNVLSPLPSQAMD 41  
 51 VFHLEGMTTSVMAQFNLLSSTMDQMSSRAASAPYTPEHAASVPTHSPYA 100  
 42 DLML...SPDDIEQWFTEDPGPDEAPRMPEAAPVAPAPAPAPTA..APAP 87  
 101 QPSTFTDMSAPVPIPSNTDYPGPHHFEVTFQQSSTAKSATWTYSPLKK 150  
 88 APSWPL.....SSSVPSQKTYQGSYGFRGLGFLHSGTAKSVTCTYSPALNK 132  
 151 LYCQIAKTCPIQIKVSTPPPGTAIRAMPVYKKAHVTDVVKRCPNHEL 200  
 133 MFCQLAKTCPVQLWVDSTPPPGTRVRAMAIYKQSQHMTVEVVRCPHHE.. 180  
 201 RDNFEGQSAPASHLIRVEGNNLSQYVDDPVTGRQSVVVPYEPQVGTFT 250  
 181 RCDSDGLAPPQHILIRVEGNLRVEYLLDRNTFRHSVVVPYEPPEVSGDCT 230  
 251 TILYNFMCNSSCVGMNRRPILIIITLEMRDQVGLGRSFEGRICACPR 300  
 231 TIHYNFMCNSSCMGMNRRPILTIITLEDSSGNLLGRNSFEVRVCACPR 280  
 301 DRKADEHYREQQALNESSAKNGAASKRAFQSPPAVPALGAGVKRRHG 350  
 281 DRRTEENLRKKGEPHHELP..PGSTKRALPNNTSSSPQ....PKKKPL 323  
 351 DEDTYYLQVRGRENFEILMKLKESLELMELVPQPLVDSYRQQQLQRP 400  
 324 DGEYFTLQIRGRERFEMFRELNEALELKDAQAGKEPGGSRAHSSHLKSK 373  
 401 HLQPPSYGPVLSPMNKVHGGMNKLPSVNQLVGQPPPHSSAATPNLGPVGP 450  
 374 GQSTSRHKKLMFKTEGPDSD 393  
 451 GMLNNHGHAVPANGEMSSSSHAQSMVSGSHCTPPPPYHADPSLVSFLTGL 500  
 501 GCPNCIEYFTSQGLQSIYHLQNLTIEDLGALKIPEQYRMTIWRGLQDLKQ 550  
 551 GHDYSTAQQLLRSSNAATISIGSGSELQRQRVMEAVHFRVRHTITIPNRG 600  
 601 GPGGGPDEWADFGFDLPDCKARKQPIKEEFTEAEIH 636

FIG.12

## FIG.13



20/36

sr-p70d-lmr32 CG ACCTTCCCCA GTCAAGCCGG GGGAAATATG 32  
 sr-p70a-ht29 CG ACCTTCCCCA GTCAAGCCGG GGGAAATATG 150

AGGTGGTGGG CGGAACGGAT TCCAGCATGG ACGTCTTCCA CCTGGAGGGC 82  
 AGGTGGTGGG CGGAACGGAT TCCAGCATGG ACGTCTTCCA CCTGGAGGGC 200

ATGACTACAT CTGTCATGCA TCCTCGGCTC CTGCCTCACT AGCTGCGGAG 132  
 ATGACTACAT CTGTCAT... .. 217

CCTCTCCCGC TCGGTCCACG CTGCCGGGGC GCCACGACCG TGACCTTCC 182  
 .....

CCTCGGGCCG CCCAGATCCA TGCCTCGTCC CACGGGACAC CAGTTCCCTG 232  
 .....

GCGTGTGCAG ACCCCCCGGC GCCTACCATG CTGTACGTGC GTGACCCCGC 282  
 .....

ACGGCACCTC GCCACGGCCC AGTTCAATCT GCTGAGCAGC ACCATGGACC 332  
 .....GGCCC AGTTCAATCT GCTGAGCAGC ACCATGGACC 252

AGATGAGCAG CCGCGCGGCC TCGGCCAGCC CCTACACCCC AGAGCACGCC 382  
 AGATGAGCAG CCGCGCGGCC TCGGCCAGCC CCTACACCCC AGAGCACGCC 302

GCCAGCGTGC CCACCCAATC GCCCTACGCA CAACCCAGCT CCACCTTCGA 432  
 GCCAGCGTGC CCACCCAATC GCCCTACGCA CAACCCAGCT CCACCTTCGA 352

CACCATGTCG CCGGCGCCTG TCATCCCCCTC CAACACCGAC TACCCCGGAC 482  
 CACCATGTCG CCGGCGCCTG TCATCCCCCTC CAACACCGAC TACCCCGGAC 402

CCCACCACCTT TGAGGTCACT TTCCAGCAGT CCAGCACGSC CAAGTCAGCC 532  
 CCCACCACCTT TGAGGTCACT TTCCAGCAGT CCAGCACGSC CAAGTCAGCC 452

ACCTGGACGT ACTCCCCGCT CTTGAAG  
 ACCTGGACGT ACTCCCCGCT CTTGAAG

FIG. 14

sr-p7/0a	T A A C G C C C G G C C G C C T A C T C C C C G G G C C T C C C C G G C C C A	50
sr-p7/0f	-	0
sr-p7/0d	-	0
sr-p7/0e	-	0
sr-p7/0b	-	0
sr-p7/0a	T A T A C C C C T A G G G G C C G G G C A G C C C G C C T C C C C G C C G C C A	100
sr-p7/0f	-	0
sr-p7/0d	-	0
sr-p7/0e	-	0
sr-p7/0b	-	0
sr-p7/0a	C C C G C C C G A G C C T C G C G C C C C G C G A G G G G A C G C A G C G A A C C G G G C C	150
sr-p7/0f	-	0
sr-p7/0d	-	0
sr-p7/0e	-	0
sr-p7/0b	-	0
sr-p7/0a	C C G C C C A G C C A G C C G G G A C G G A C C C A T G C C C G G G G C T G C G A C C G G C T	200
sr-p7/0f	-	0
sr-p7/0d	-	20
sr-p7/0e	-	0
sr-p7/0b	-	0
sr-p7/0a	G C A G A G C G A G T C C C C T C G G A G G C C G G C G T G G G G A A G A T G G C C C A G T C C A	250
sr-p7/0f	G C A G -	24
sr-p7/0d	-	0
sr-p7/0e	-	0
sr-p7/0b	-	13

FIG. 15

sr-p70a	CGCCACCTCCCTTGATGGGGGCAACCACTTTTGAGCACCTCTGGAGCTCTT	303
sr-p70b	- - - - -	24
sr-p70f	- - - - -	0
sr-p70d	- - - - -	0
sr-p70e	- - - - -	0
sr-p70b	CGCCACCTCCCTTGATGGGGGCAACCACTTTTGAGCACCTCTGGAGCTCTT	63
sr-p70a	CTGGAAACCAGACAGCACTTACTTCCGACCTTCCCAAGTCAAGCCGGGGGAA	350
sr-p70f	- GGAACCAAGACAGCACTTACTTCCGACCTTCCCAAGTCAAGCCGGGGGAA	79
sr-p70d	- - - - -	0
sr-p70e	- - - - -	0
sr-p70b	CTGGAAACCAGACAGCACTTACTTCCGACCTTCCCAAGTCAAGCCGGGGGAA	113
sr-p70a	TAAATGAGGTGGTGGGGGGAACGGATTCCAGCATGGACGTCTTCCACCTGG	400
sr-p70f	TAAATGAGGTGGTGGGGGGAACGGATTCCAGCATGGACGTCTTCCACCTGG	122
sr-p70d	- - - - -	33
sr-p70e	- - - - -	33
sr-p70b	TAAATGAGGTGGTGGGGGGAACGGATTCCAGCATGGACGTCTTCCACCTGG	163
sr-p70a	AGGGCATGACTACATCTGTTCATGGCCCAAGTTCAATCTGCTGAGCAGCACCC	450
sr-p70f	AGGGCATGACTACATCTGTTCATGGCCCAAGTTCAATCTGCTGAGCAGCACCC	172
sr-p70d	- - - - -	66
sr-p70e	- - - - -	66
sr-p70b	AGGGCATGACTACATCTGTTCATGGCCCAAGTTCAATCTGCTGAGCAGCACCC	213
sr-p70a	ATGGACCAAGATGAGCAGCCGGCCGGCCCTCGGCCAGCCCCCTACACCCAGAGA	500
sr-p70f	ATGGACCAAGATGAGCAGCCGGCCGGCCCTCGGCCAGCCCCCTACACCCAGAGA	222
sr-p70d	ATGGACCAAGATGAGCAGCCGGCCGGCCCTCGGCCAGCCCCCTACACCCAGAGA	116
sr-p70e	ATGGACCAAGATGAGCAGCCGGCCGGCCCTCGGCCAGCCCCCTACACCCAGAGA	116
sr-p70b	ATGGACCAAGATGAGCAGCCGGCCGGCCCTCGGCCAGCCCCCTACACCCAGAGA	263

sr-p70a	GCACGCGCGCACGCGTGCCCA	CCCACCCACCTCGGCCCTACGCGCACAA	CCAGCTCCA	550
sr-p70f	GCACGCGCGCACGCGTGCCCA	CCCACCCACCTCGGCCCTACGCGCACAA	CCAGCTCCA	572
sr-p70d	GCACGCGCGCACGCGTGCCCA	CCCACCCACCTCGGCCCTACGCGCACAA	CCAGCTCCA	166
sr-p70e	GCACGCGCGCACGCGTGCCCA	CCCACCCACCTCGGCCCTACGCGCACAA	CCAGCTCCA	166
sr-p70b	GCACGCGCGCACGCGTGCCCA	CCCACCCACCTCGGCCCTACGCGCACAA	CCAGCTCCA	313
sr-p70a	CCCTTCGACACCATGTCTGCGCGCGCGCGCTGTCTCAT	CCCCCTCCACACACCGACTAC		600
sr-p70f	CCCTTCGACACCATGTCTGCGCGCGCGCGCTGTCTCAT	CCCCCTCCACACACCGACTAC		322
sr-p70d	CCCTTCGACACCATGTCTGCGCGCGCGCGCTGTCTCAT	CCCCCTCCACACACCGACTAC		216
sr-p70e	CCCTTCGACACCATGTCTGCGCGCGCGCGCTGTCTCAT	CCCCCTCCACACACCGACTAC		216
sr-p70b	CCCTTCGACACCATGTCTGCGCGCGCGCGCTGTCTCAT	CCCCCTCCACACACCGACTAC		363
sr-p70a	CCCGGACCCCAACCATTTTGAGGGTCACTTTCCAGCAGTCCAGCACGGGCCAA			650
sr-p70f	CCCGGACCCCAACCATTTTGAGGGTCACTTTCCAGCAGTCCAGCACGGGCCAA			372
sr-p70d	CCCGGACCCCAACCATTTTGAGGGTCACTTTCCAGCAGTCCAGCACGGGCCAA			266
sr-p70e	CCCGGACCCCAACCATTTTGAGGGTCACTTTCCAGCAGTCCAGCACGGGCCAA			266
sr-p70b	CCCGGACCCCAACCATTTTGAGGGTCACTTTCCAGCAGTCCAGCACGGGCCAA			413
sr-p70a	GTCAGCCACCTGGACGTA	CTCCCGCTCTTGAAAGAACTCTACTGCCCCAGA		700
sr-p70f	GTCAGCCACCTGGACGTA	CTCCCGCTCTTGAAAGAACTCTACTGCCCCAGA		422
sr-p70d	GTCAGCCACCTGGACGTA	CTCCCGCTCTTGAAAGAACTCTACTGCCCCAGA		316
sr-p70e	GTCAGCCACCTGGACGTA	CTCCCGCTCTTGAAAGAACTCTACTGCCCCAGA		316
sr-p70b	GTCAGCCACCTGGACGTA	CTCCCGCTCTTGAAAGAACTCTACTGCCCCAGA		463
sr-p70a	TCGCCCCAAGACATGCCCCCATCTCCAGATCAAGGTGTCTCCAC	CCCGGCCACCCCCCA		750
sr-p70f	TCGCCCCAAGACATGCCCCCATCTCCAGATCAAGGTGTCTCCAC	CCCGGCCACCCCCCA		472
sr-p70d	TCGCCCCAAGACATGCCCCCATCTCCAGATCAAGGTGTCTCCAC	CCCGGCCACCCCCCA		366
sr-p70e	TCGCCCCAAGACATGCCCCCATCTCCAGATCAAGGTGTCTCCAC	CCCGGCCACCCCCCA		366
sr-p70b	TCGCCCCAAGACATGCCCCCATCTCCAGATCAAGGTGTCTCCAC	CCCGGCCACCCCCCA		513

FIG.15 cont.

sr-p70a 800  
sr-p70f 522  
sr-p70d 416  
sr-p70e 416  
sr-p70b 563

GGCACTGCGCATCCGGGCCATGCTGTGTTTACAAAGAAAGCGGAGCAACGTGAC  
GGCACTGCGCATCCGGGCCATGCTGTGTTTACAAAGAAAGCGGAGCAACGTGAC  
GGCACTGCGCATCCGGGCCATGCTGTGTTTACAAAGAAAGCGGAGCAACGTGAC  
GGCACTGCGCATCCGGGCCATGCTGTGTTTACAAAGAAAGCGGAGCAACGTGAC

sr-p70a 850  
sr-p70f 572  
sr-p70d 466  
sr-p70e 466  
sr-p70b 613

CGACGTCGTGAAACGCTGCCCAACCAACGAGCTCGGGAGGACCTTCAACG  
CGACGTCGTGAAACGCTGCCCAACCAACGAGCTCGGGAGGACCTTCAACG  
CGACGTCGTGAAACGCTGCCCAACCAACGAGCTCGGGAGGACCTTCAACG  
CGACGTCGTGAAACGCTGCCCAACCAACGAGCTCGGGAGGACCTTCAACG

sr-p70a 900  
sr-p70f 622  
sr-p70d 516  
sr-p70e 516  
sr-p70b 663

AAGGACAGTCTGCTCCAGCCAGCCACCTCATCTCCGCTGGAGGCAATAAT  
AAGGACAGTCTGCTCCAGCCAGCCACCTCATCTCCGCTGGAGGCAATAAT  
AAGGACAGTCTGCTCCAGCCAGCCACCTCATCTCCGCTGGAGGCAATAAT  
AAGGACAGTCTGCTCCAGCCAGCCACCTCATCTCCGCTGGAGGCAATAAT

sr-p70a 950  
sr-p70f 672  
sr-p70d 566  
sr-p70e 566  
sr-p70b 713

CTCTCGCAGTATGTGGATGACCTCTGTCAACCGGCAAGGCAGAGCGTGGT  
CTCTCGCAGTATGTGGATGACCTCTGTCAACCGGCAAGGCAGAGCGTGGT  
CTCTCGCAGTATGTGGATGACCTCTGTCAACCGGCAAGGCAGAGCGTGGT  
CTCTCGCAGTATGTGGATGACCTCTGTCAACCGGCAAGGCAGAGCGTGGT

sr-p70a 1000  
sr-p70f 722  
sr-p70d 616  
sr-p70e 616  
sr-p70b 763

GCCCTATGAGCCACCAACAGGTGGGGACGGGAATTCACCAACCATCCTGTACA  
GCCCTATGAGCCACCAACAGGTGGGGACGGGAATTCACCAACCATCCTGTACA  
GCCCTATGAGCCACCAACAGGTGGGGACGGGAATTCACCAACCATCCTGTACA  
GCCCTATGAGCCACCAACAGGTGGGGACGGGAATTCACCAACCATCCTGTACA

FIG.15 cont.

sr-p70a A C T T C A T G T G T A A C A G C A G C T G T G T A G G G G C C A T G A A C C G G C G G C C C C A T C 1050  
 sr-p70f A C T T C A T G T G T A A C A G C A G C T G T G T A G G G G C C A T G A A C C G G C G G C C C A T C 772  
 sr-p70d A C T T C A T G T G T A A C A G C A G C T G T G T A G G G G C C A T G A A C C G G C G G C C C A T C 666  
 sr-p70e A C T T C A T G T G T A A C A G C A G C T G T G T A G G G G C C A T G A A C C G G C G G C C C A T C 666  
 sr-p70b A C T T C A T G T G T A A C A G C A G C T G T G T A G G G G C C A T G A A C C G G C G G C C C A T C 813

sr-p70a C T C A T C A T C A T C A C C C T G G A G A T G C G G G A T G G G C A G G T G C T G G G C C G C C G 1100  
 sr-p70f C T C A T C A T C A T C A C C C T G G A G A T G C G G G A T G G G C A G G T G C T G G G C C G C C G 822  
 sr-p70d C T C A T C A T C A T C A C C C T G G A G A T G C G G G A T G G G C A G G T G C T G G G C C G C C G 716  
 sr-p70e C T C A T C A T C A T C A C C C T G G A G A T G C G G G A T G G G C A G G T G C T G G G C C G C C G 716  
 sr-p70b C T C A T C A T C A T C A C C C T G G A G A T G C G G G A T G G G C A G G T G C T G G G C C G C C G 863

sr-p70a G T C C T T T G A G G G C C G C A T C T G C G C C C T G T C C T G G C C C G C G A C C G A A A A G C T G 1150  
 sr-p70f G T C C T T T G A G G G C C G C A T C T G C G C C C T G T C C T G G C C C G C G A C C G A A A A G C T G 872  
 sr-p70d G T C C T T T G A G G G C C G C A T C T G C G C C C T G T C C T G G C C C G C G A C C G A A A A G C T G 766  
 sr-p70e G T C C T T T G A G G G C C G C A T C T G C G C C C T G T C C T G G C C C G C G A C C G A A A A G C T G 766  
 sr-p70b G T C C T T T G A G G G C C G C A T C T G C G C C C T G T C C T G G C C C G C G A C C G A A A A G C T G 913

sr-p70a A T G A G G A C C A C T A C C G G G A G C A G C A G G C C C T G A A C G A G A G C T C C G C C A A G 1200  
 sr-p70f A T G A G G A C C A C T A C C G G G A G C A G C A G G C C C T G A A C G A G A G C T C C G C C A A G 922  
 sr-p70d A T G A G G A C C A C T A C C G G G A G C A G C A G G C C C T G A A C G A G A G C T C C G C C A A G 816  
 sr-p70e A T G A G G A C C A C T A C C G G G A G C A G C A G G C C C T G A A C G A G A G C T C C G C C A A G 816  
 sr-p70b A T G A G G A C C A C T A C C G G G A G C A G C A G G C C C T G A A C G A G A G C T C C G C C A A G 963

sr-p70a A A C G G G G C C G C C A G C A A G C G T G C C T T C A A G C A G A G C C C C C C T G C C G T C C C 1250  
 sr-p70f A A C G G G G C C G C C A G C A A G C G T G C C T T C A A G C A G A G C C C C C C T G C C G T C C C 972  
 sr-p70d A A C G G G G C C G C C A G C A A G C G T G C C T T C A A G C A G A G C C C C C C T G C C G T C C C 866  
 sr-p70e A A C G G G G C C G C C A G C A A G C G T G C C T T C A A G C A G A G C C C C C C T G C C G T C C C 866  
 sr-p70b A A C G G G G C C G C C A G C A A G C G T G C C T T C A A G C A G A G C C C C C C T G C C G T C C C 1013

FIG. 15 cont.

sr-p70a	CGTCTACGGGCGGGTCCCTCTCGCCCATGAAACAAGGTGCCACGGGGGCGCATG	1500
sr-p70f	CGTCTACGGGCGGGTCCCTCTCGCCCATGAAACAAGGTGCCACGGGGGCGCATG	1222
sr-p70d	CGTCTACGGGCGGGTCCCTCTCGCCCATGAAACAAGGTGCCACGGGGGCGCATG	1116
sr-p70e	-----	1049
sr-p70b	CGTCTACGGGCGGGTCCCTCTCGCCCATGAAACAAGGTGCCACGGGGGCGCATG	1263

FIG. 15 cont.

sr-p70a	A A C A A G C T G C C C C T C C C G T C A A C C A G C T G G T G G G C C A G C C C T C C C C G C A C A G	1550
sr-p70f	A A C A A G C T G C C C C T C C C G T C A A C C A G C T G G T G G G C C A G C C T C C C C G C A C A G	1272
sr-p70d	A A C A A G C T G C C C C T C C C G T C A A C C A G C T G G T G G G C C A G C C T C C C C G C A C A G	1166
sr-p70e	-----	1049
sr-p70b	A A C A A G C T G C C C C T C C C G T C A A C C A G C T G G T G G G C C A G C C T C C C C G C A C A G	1313
sr-p70a	T T C G G C A G C T A C A C C C C A A C C T G G G G C C C C G T G G C C C C G G G A T G C T C A A C A	1600
sr-p70f	T T C G G C A G C T A C A C C C C A A C C T G G G G C C C C G T G G C C C C G G A T G C T C A A C A	1322
sr-p70d	T T C G G C A G C T A C A C C C C A A C C T G G G G C C C C G T G G C C C C G G A T G C T C A A C A	1216
sr-p70e	-----	1067
sr-p70b	T T C G G C A G C T A C A C C C C A A C C T G G G G C C C C G T G G C C C C G G G A T G C T C A A C A	1363
sr-p70a	A C C A T G G G C C A C G C A G T G C C A G G C C A A C G G C G G A G A T G A G C A G C A G C C A C A G C	1650
sr-p70f	A C C A T G G G C C A C G C A G T G C C A G G C C A A C G G C G G A G A T G A G C A G C A G C C A C A G C	1372
sr-p70d	A C C A T G G G C C A C G C A G T G C C A G G C C A A C G G C G G A G A T G A G C A G C A G C C A C A G C	1266
sr-p70e	A C C A T G G G C C A C G C A G T G C C A G G C C A A C G G C G G A G A T G A G C A G C A G C C A C A G C	1117
sr-p70b	A C C A T G G G C C A C G C A G T G C C A G G C C A A C G G C G G A G A T G A G C A G C A G C C A C A G C	1413
sr-p70a	G C C C A G T C C A T G G T C T C G G G G T C C C A C T G C A C T C C G C C A C C C C C C C T A C C A	1700
sr-p70f	G C C C A G T C C A T G G T C T C G G G G T C C C A C T G C A C T C C G C C A C C C C C C T A C C A	1422
sr-p70d	G C C C A G T C C A T G G T C T C G G G G T C C C A C T G C A C T C C G C C A C C C C C C T A C C A	1316
sr-p70e	G C C C A G T C C A T G G T C T C G G G G T C C C A C T G C A C T C C G C C A C C C C C C T A C C A	1167
sr-p70b	G C C C A G T C C A T G G T C T C G G G G T C C C A C T G C A C T C C G C C A C C C C C C T A C C A	1463
sr-p70a	C G C C G A C C C C A G C C T C G T C A G T T T T T A A C A G A G A T T G G G G T G T C C A A A C T	1750
sr-p70f	C G C C G A C C C C A G C C T C G T C A G T T T T T A A C A G A G A T T G G G G T G T C C A A A C T	1472
sr-p70d	C G C C G A C C C C A G C C T C G T C A G T T T T T A A C A G A G A T T G G G G T G T C C A A A C T	1366
sr-p70e	C G C C G A C C C C A G C C T C G T C A G T T T T T A A C A G A G A T T G G G G T G T C C A A A C T	1186
sr-p70b	C G C C G A C C C C A G C C T C G T C A G T T T T T A A C A G A G A T T G G G G T G T C C A A A C T	1482

FIG.15 cont.

[illegible]

sr-p70a	A A C C T G A C C A T T G A G G A C C T G G G G G C C C T G A A G A T C C C C G G A G C A G T A C C G	1050
sr-p70f	A A C C T G A C C A T T G A G G A C C T G G G G G C C C T G A A G A T C C C C G G A G C A G T A C C G	1572
sr-p70d	A A C C T G A C C A T T G A G G A C C T G G G G G C C C T G A A G A T C C C C G G A G C A G T A C C G	1466
sr-p70e	- - - - - - - - - - A G G A C C T G G G G G C C C T G A A G A T C C C C G G A G C A G T A C C G	1223
sr-p70b	- - - - - - - - - - A G G A C C T G G G G G C C C T G A A G A T C C C C G G A G C A G T A C C G	1519

sr-p70a	CATGACCATCTGGCGGGGCTTGCAGGACCTGTGAAGCAGGGCCACGGACTACA	1900
sr-p70f	CATGACCATCTGGCGGGGCTTGCAGGACCTGTGAAGCAGGGCCACGGACTACA	1622
sr-p70d	CATGACCATCTGGCGGGGCTTGCAGGACCTGTGAAGCAGGGCCACGGACTACA	1516
sr-p70e	CATGACCATCTGGCGGGGCTTGCAGGACCTGTGAAGCAGGGCCACGGACTACA	1273
sr-p70b	CATGACCATCTGGCGGGGCTTGCAGGACCTGTGAAGCAGGGCCACGGACTACA	1569

sr-p70a	G C A C G C G C A G A G C T G C T C C G C T C T A G C A A C G C G G C C A C C A T C T C C A T C	1950
sr-p70f	G C A C G C G C A G A G C T G C T C C G C T C T A G C A A C G C G C C A C C A T C T C C A T C	1672
sr-p70d	G C A C G C G C A G A G C T G C T C C G C T C T A G C A A C G C G C C A C C A T C T C C A T C	1566
sr-p70e	G C A C G C G C A G A G C T G C T C C G C T C T A G C A A C G C G C C A C C A T C T C C A T C	1323
sr-p70b	G C A C G C G C A G A G C T G C T C C G C T C T A G C A A C G C G C C A C C A T C T C C A T C	1619

sr-p70a	GGCGGCTCAGGGGAAC	TGCAGCGCCAGCGGGT	CATGGAGGCCGTC	CACTT	2000
sr-p70f	GGCGCTCAGGGAAC	TGCAGCGCCAGCGGGT	CATGGAGGCCGTC	CACTT	1722
sr-p70e	GGCGCTCAGGGAAC	TGCAGCGCCAGCGGGT	CATGGAGGCCGTC	CACTT	1616
sr-p70d	GGCGCTCAGGGAAC	TGCAGCGCCAGCGGGT	CATGGAGGCCGTC	CACTT	1373
sr-p70b	GGCGGCTCAGGGGAAC	TGCAGCGCCAGCGGGT	CATGGAGGCCGTC	CACTT	1669

FIG. 15 cont.

sr-p70a	CCGCGGTGCGCCACACCATCAACCATCCCCAAACCGCGCGCCAGGCGGCGG	2050
sr-p70f	CCGCGGTGCGCCACACCATCAACCATCCCCAAACCGCGCGCCAGGCGGCGG	1772
sr-p70d	CCGCGGTGCGCCACACCATCAACCATCCCCAAACCGCGCGCCAGGCGGCGG	1666
sr-p70e	CCGCGGTGCGCCACACCATCAACCATCCCCAAACCGCGCGCCAGGCGGCGG	1423
sr-p70b	CCGCGGTGCGCCACACCATCAACCATCCCCAAACCGCGCGCCAGGCGGCGG	1719
sr-p70a	GCCCTGACGAGTGGGCGGGACTTCGGGCTTCGACCTGCCCCGACTGCAAGGCC	2100
sr-p70f	GCCCTGACGAGTGGGCGGGACTTCGGGCTTCGACCTGCCCCGACTGCAAGGCC	1822
sr-p70d	GCCCTGACGAGTGGGCGGGACTTCGGGCTTCGACCTGCCCCGACTGCAAGGCC	1716
sr-p70e	GCCCTGACGAGTGGGCGGGACTTCGGGCTTCGACCTGCCCCGACTGCAAGGCC	1473
sr-p70b	GCCCTGACGAGTGGGCGGGACTTCGGGCTTCGACCTGCCCCGACTGCAAGGCC	1769
sr-p70a	CGCAAGCAGGCCCATCAAGGAGGAGTTTCA CGGAGGCCGAGATCCACTGA	2150
sr-p70f	CGCAAGCAGGCCCATCAAGGAGGAGTTTCA CGGAGGCCGAGATCCACTGA	1870
sr-p70d	CGCAAGCAGGCCCATCAAGGAGGAGTTTCA CGGAGGCCGAGATCCACTGA	1764
sr-p70e	CGCAAGCAGGCCCATCAAGGAGGAGTTTCA CGGAGGCCGAGATCCACTGA	1521
sr-p70b	CGCAAGCAGGCCCATCAAGGAGGAGTTTCA CGGAGGCCGAGATCCACTGA	1817
sr-p70a	GCCCTCGCTGCTGAGCTGCGCCACCGCCAGAGACCCAGCTGCTC	2200
sr-p70f	GCCCTCGCTGCTGAGCTGCGCCACCGCCAGAGACCCAGCTGCTC	1870
sr-p70d	GCCCTCGCTGCTGAGCTGCGCCACCGCCAGAGACCCAGCTGCTC	1764
sr-p70e	GCCCTCGCTGCTGAGCTGCGCCACCGCCAGAGACCCAGCTGCTC	1521
sr-p70b	GCCCTCGCTGCTGAGCTGCGCCACCGCCAGAGACCCAGCTGCTC	1817
sr-p70a	CCCTCTCCTCCTGCTGTGTCCAAACTGCTCAGGAGCAGGACCTTCGG	2250
sr-p70f	CCCTCTCCTCCTGCTGTGTCCAAACTGCTCAGGAGCAGGACCTTCGG	1870
sr-p70d	CCCTCTCCTCCTGCTGTGTCCAAACTGCTCAGGAGCAGGACCTTCGG	1764
sr-p70e	CCCTCTCCTCCTGCTGTGTCCAAACTGCTCAGGAGCAGGACCTTCGG	1521
sr-p70b	CCCTCTCCTCCTGCTGTGTCCAAACTGCTCAGGAGCAGGACCTTCGG	1817

FIG.15 cont.

Accession	Gene	Accession	Gene	Accession	Gene
sr-p7a	GCCCCAGGAAAGGCCCCAGCCACCGAAGCCCGCTGTGTGGACAGCCTGAGTCA	sr-p7a	GCCCCAGGAAAGGCCCCAGCCACCGAAGCCCGCTGTGTGGACAGCCTGAGTCA	sr-p7a	GCCCCAGGAAAGGCCCCAGCCACCGAAGCCCGCTGTGTGGACAGCCTGAGTCA
sr-p7b	-----	sr-p7b	-----	sr-p7b	-----
sr-p7c	-----	sr-p7c	-----	sr-p7c	-----
sr-p7d	-----	sr-p7d	-----	sr-p7d	-----
sr-p7e	-----	sr-p7e	-----	sr-p7e	-----
sr-p7f	-----	sr-p7f	-----	sr-p7f	-----
sr-p7g	-----	sr-p7g	-----	sr-p7g	-----
sr-p7h	-----	sr-p7h	-----	sr-p7h	-----
sr-p7i	-----	sr-p7i	-----	sr-p7i	-----
sr-p7j	-----	sr-p7j	-----	sr-p7j	-----
sr-p7k	-----	sr-p7k	-----	sr-p7k	-----
sr-p7l	-----	sr-p7l	-----	sr-p7l	-----
sr-p7m	-----	sr-p7m	-----	sr-p7m	-----
sr-p7n	-----	sr-p7n	-----	sr-p7n	-----
sr-p7o	-----	sr-p7o	-----	sr-p7o	-----
sr-p7p	-----	sr-p7p	-----	sr-p7p	-----
sr-p7q	-----	sr-p7q	-----	sr-p7q	-----
sr-p7r	-----	sr-p7r	-----	sr-p7r	-----
sr-p7s	-----	sr-p7s	-----	sr-p7s	-----
sr-p7t	-----	sr-p7t	-----	sr-p7t	-----
sr-p7u	-----	sr-p7u	-----	sr-p7u	-----
sr-p7v	-----	sr-p7v	-----	sr-p7v	-----
sr-p7w	-----	sr-p7w	-----	sr-p7w	-----
sr-p7x	-----	sr-p7x	-----	sr-p7x	-----
sr-p7y	-----	sr-p7y	-----	sr-p7y	-----
sr-p7z	-----	sr-p7z	-----	sr-p7z	-----

sr-p70a C C T G C A G A A C C  
sr-p70f - - - - -  
sr-p70d - - - - -  
sr-p70e - - - - -  
sr-p70b - - - - -

2361  
1870  
1764  
1521  
1817

FIG. 15 cont.

sr-p70a_	MAQSTATSPDGGTTPEHLWSSLEPDSYFDLPQSSRGNNVVGGTDSMD	50
sr-p70f_	-----	2
sr-p70d_	-----	1
sr-p70b_	MAQSTATSPDGGTTPEHLWSSLEPDSYFDLPQSSRGNNVVGGTDSMD	50
sr-p70e_	-----	1
sr-p70a_	VFHLEGMTTSVMAQPNLLSSTMDQMSRRAASASPYTPEHAASVPTTHSPYA	100
sr-p70f_	VFHLEGMTTSVMAQPNLLSSTMDQMSRRAASASPYTPEHAASVPTTHSPYA	52
sr-p70d_	LYVGDPAARHLATQVFNLLSSTMDQMSRRAASASPYTPEHAASVPTTHSPYA	51
sr-p70b_	VFHLEGMTTSVMAQPNLLSSTMDQMSRRAASASPYTPEHAASVPTTHSPYA	100
sr-p70e_	LYVGDPAARHLATQVFNLLSSTMDQMSRRAASASPYTPEHAASVPTTHSPYA	51
sr-p70a_	QPSSTFTDMSAPVIPSNTDYPGPHHFEVTFQSSSTAKSATMTYSPLLLK	150
sr-p70f_	QPSSTFTDMSAPVIPSNTDYPGPHHFEVTFQSSSTAKSATMTYSPLLLK	102
sr-p70d_	QPSSTFTDMSAPVIPSNTDYPGPHHFEVTFQSSSTAKSATMTYSPLLLK	101
sr-p70b_	QPSSTFTDMSAPVIPSNTDYPGPHHFEVTFQSSSTAKSATMTYSPLLLK	150
sr-p70e_	QPSSTFTDMSAPVIPSNTDYPGPHHFEVTFQSSSTAKSATMTYSPLLLK	101
sr-p70a_	LYCQIAKTCPIQIKVSTPPPGTAIRAMPVYKKAHEVTDVVKRCPNHEL	200
sr-p70f_	LYCQIAKTCPIQIKVSTPPPGTAIRAMPVYKKAHEVTDVVKRCPNHEL	152
sr-p70d_	LYCQIAKTCPIQIKVSTPPPGTAIRAMPVYKKAHEVTDVVKRCPNHEL	151
sr-p70b_	LYCQIAKTCPIQIKVSTPPPGTAIRAMPVYKKAHEVTDVVKRCPNHEL	200
sr-p70e_	LYCQIAKTCPIQIKVSTPPPGTAIRAMPVYKKAHEVTDVVKRCPNHEL	151
sr-p70a_	RDFNEGQSAASHLIRVEGNNLSQYVDDPVTGRQSVVVVPEPPQVGTFT	250
sr-p70f_	RDFNEGQSAASHLIRVEGNNLSQYVDDPVTGRQSVVVVPEPPQVGTFT	202
sr-p70d_	RDFNEGQSAASHLIRVEGNNLSQYVDDPVTGRQSVVVVPEPPQVGTFT	201
sr-p70b_	RDFNEGQSAASHLIRVEGNNLSQYVDDPVTGRQSVVVVPEPPQVGTFT	250
sr-p70e_	RDFNEGQSAASHLIRVEGNNLSQYVDDPVTGRQSVVVVPEPPQVGTFT	201

sr-p70a-	T I L Y N F M C N S S C V G G M N R R P I L I I T L E M R D G Q V L G R R S F E G R I C A C P G R	300
sr-p70f-	T I L Y N F M C N S S C V G G M N R R P I L I I T L E M R D G Q V L G R R S F E G R I C A C P G R	352
sr-p70d-	T I L Y N F M C N S S C V G G M N R R P I L I I T L E M R D G Q V L G R R S F E G R I C A C P G R	351
sr-p70b-	T I L Y N F M C N S S C V G G M N R R P I L I I T L E M R D G Q V L G R R S F E G R I C A C P G R	300
sr-p70e-	T I L Y N F M C N S S C V G G M N R R P I L I I T L E M R D G Q V L G R R S F E G R I C A C P G R	251
sr-p70a-	DRK A D E D H Y R E Q Q A L N E S S A K N G A A S K R A F K Q S P P A V P A L G A G V K K R R H G	350
sr-p70f-	DRK A D E D H Y R E Q Q A L N E S S A K N G A A S K R A F K Q S P P A V P A L G A G V K K R R H G	302
sr-p70d-	DRK A D E D H Y R E Q Q A L N E S S A K N G A A S K R A F K Q S P P A V P A L G A G V K K R R H G	301
sr-p70b-	DRK A D E D H Y R E Q Q A L N E S S A K N G A A S K R A F K Q S P P A V P A L G A G V K K R R H G	350
sr-p70e-	DRK A D E D H Y R E Q Q A L N E S S A K N G A A S K R A F K Q S P P A V P A L G A G V K K R R H G	301
sr-p70a-	D E D T Y Y L Q V R G R E N F E I L M K L K E S L E L M E L V P Q P L V D S Y R Q Q Q L L Q R P S	400
sr-p70f-	D E D T Y Y L Q V R G R E N F E I L M K L K E S L E L M E L V P Q P L V D S Y R Q Q Q L L Q R P S	352
sr-p70d-	D E D T Y Y L Q V R G R E N F E I L M K L K E S L E L M E L V P Q P L V D S Y R Q Q Q L L Q R P S	351
sr-p70b-	D E D T Y Y L Q V R G R E N F E I L M K L K E S L E L M E L V P Q P L V D S Y R Q Q Q L L Q R P S	400
sr-p70e-	D E D T Y Y L Q V R G R E N F E I L M K L K E S L E L M E L V P Q P L V D S Y R Q Q Q L L Q R P P	351
sr-p70a-	H L Q P P S Y G P V L S P M N K V H G M N K L P S V N Q L V G Q P P P H S S A A T P N L G P V G P	450
sr-p70f-	H L Q P P S Y G P V L S P M N K V H G M N K L P S V N Q L V G Q P P P H S S A A T P N L G P V G P	402
sr-p70d-	H L Q P P S Y G P V L S P M N K V H G M N K L P S V N Q L V G Q P P P H S S A A T P N L G P V G P	401
sr-p70b-	H L Q P P S Y G P V L S P M N K V H G M N K L P S V N Q L V G Q P P P H S S A A T P N L G P V G P	450
sr-p70e-	R D A Q Q P W P - - - - - R S A S Q R R D E Q Q P Q R P V - - - - -	375
sr-p70a-	G M L N N H G H A V P A N G E M S S S H S A Q S M V S G S H C T P P P P Y H A D P S L V S F L T G L	500
sr-p70f-	G M L N N H G H A V P A N G E M S S S H S A Q S M V S G S H C T P P P P Y H A D P S L V S F L T G L	452
sr-p70d-	G M L N N H G H A V P A N G E M S S S H S A Q S M V S G S H C T P P P P Y H A D P S L V S F L T G L	451
sr-p70b-	G M L N N H G H A V P A N G E M S S S H S A Q S M V S G S H C T P P P P Y H A D P S L V R T W G P -	499
sr-p70e-	- - - - - H G L G V P L - - - - - H S A T P L P R R R P Q P R - - - - -	395

32/36

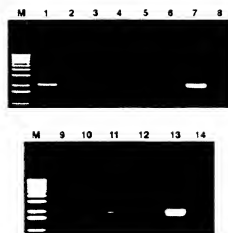
FIG. 16 cont.

sr-p70a_	GCPNCIEYFTSQGLQSIYHLQNLTIEDLGALKIPEQYRMTIWRGLQDLKQ	550
sr-p70f_	GCPNCIEYFTSQGLQSIYHLQNLTIEDLGALKIPEQYRMTIWRGLQDLKQ	502
sr-p70d_	GCPNCIEYFTSQGLQSIYHLQNLTIEDLGALKIPEQYRMTIWRGLQDLKQ	501
sr-p70b_	-----	499
sr-p70e_	-----QDLGALKIPEQYRMTIWRGLQDLKQ	420
sr-p70a_	GHDYSTAQQLLRSSNAATISIGSGGELQRQRMVMEAVHFRVRHTITIPNRG	600
sr-p70f_	GHDYSTAQQLLRSSNAATISIGSGGELQRQRMVMEAVHFRVRHTITIPNRG	552
sr-p70d_	GHDYSTAQQLLRSSNAATISIGSGGELQRQRMVMEAVHFRVRHTITIPNRG	551
sr-p70b_	-----	499
sr-p70e_	GHDYSTAQQLLRSSNAATISIGSGGELQRQRMVMEAVHFRVRHTITIPNRG	470
sr-p70a_	GPGGGPDEWADFGFDLPDCKARKQPIKEEFTAEIHH	636
sr-p70f_	GPGGGPDEWADFGFDLPDCKARKQPIKEEFTAEIHH	588
sr-p70d_	GPGGGPDEWADFGFDLPDCKARKQPIKEEFTAEIHH	587
sr-p70b_	-----	499
sr-p70e_	GPGGGPDEWADFGFDLPDCKARKQPIKEEFTAEIHH	506

G00020=30052500

1 TAAAGCCGCGGGCGGCTTACTCCCGGGGGCGCTCCCTCCCGCGCCCATATAACCCGC  
 60 CTAGGGGGCGGGGAGCCGCCCTGGCTCCCGCGCCCGCGACCCCGCGGAGCTCGGGG  
 120 CCCCGGAAGGGGAGCCGAGCGAAACCGGGGCGCGCGCCAGCCGCGGACGGACGCCGA  
 180 TCCCGGGGGCTCGGACCGCTCGAGAGCAGCTGCTCCCTTGGAGGGCGGGGTGGGGAAGATG  
 240  
 1 M  
 241 GCCCAGTCCACCGCCACCTCCCTCGATGGGGGACACACGTTTGGACGACCTCTGGAGCTCT  
 2 A Q S T A T S P D G G T T F E H L W S S  
 301 CTGGAACGAGACAGACGACCTTACTTCGAGCTTTCCTCCAGTCAGCCGGGGGAATATGAGTG  
 360 L E P D S T Y F D L P Q S S R G N E V  
 41 GTGGCGGAACGGATTCACGATGGAGCTCTTCACCTGGAGGGCATGAGCTACATCTGTG  
 420 V G G T D S M D V F H L E G M T T S V  
 61 ATGGCCGAGTTCAATCTGCTGAGCAGCACCTGGACACAGATGAGCAGCCGCGGCTCG  
 480 M A Q F N L S S T M D Q M S S R A S  
 81 GCCAGCCCTACACCCGAGAGCAGCCGCGAGTGCCACCCACTCGCCTTACGACAA  
 540 A S P Y T P E H A A S V P T H S P Y A Q  
 101 CCCAGCTCCACTTCGACACCATGTGCGCGGCGCTGTGATCCCTTCCACACGACGTAC  
 600 P S S T P D T M S P A P V I P S N T D Y  
 121 CCCGAGCCCGACCACTTTGAGGTCACTTTCAGCAGCTCAGCAGCGCCAGTCAAGACC  
 660 122 P G P H H F E V T F Q Q S S T A K S A T  
 661 TGGACGTA.....  
 141  
 142 W T

FIG.17

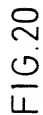
FIG. 18

M 1 2 3 4 5 6 7 8 9 10 M

FIG. 19A

M 1 2 3 4 5 6 7 8 9 10 M

FIG. 19B



PolyInker1: 0.0/HindIII,NotI,KpnI.  
PolyInker2: 2.16/XbaI,NotI,Apal.

# DECLARATION AND POWER OF ATTORNEY FOR UNITED STATES PATENT APPLICATION

  X   Original             Supplemental             Substitute

As a below-named inventor, I hereby declare that:

My residence, citizenship and post office address are given below under my name.

I believe I am an original, first and joint inventor of the subject matter which is claimed and for which a patent is sought on the invention entitled:

## Purified SR-p70 protein

the specification of which

       is attached hereto.

  X   was filed on                                  as United States  
Application Serial No.                                   
and was amended on                                  (if applicable).

       was filed on February 03, 1997 as PCT International  
Application No. PCT/FR97/00214  
and was amended under PCT Article 19 on September 02, 1997 (if applicable).

I have reviewed and understand the contents of the above-identified specification, including the claims, as amended by any amendment specifically referred to above.

I acknowledge my duty to disclose information of which I am aware which is material to the examination of this application in accordance with Section 1.56 of Title 37 of the Code of Federal Regulations.

I hereby claim foreign priority benefit under Section 119 (a) - (d) of Title 35 of the United States Code of any foreign application(s) for patent or inventor's certificate or of any PCT application(s) designating at least one country other than the United States identified below and also identify below any foreign application(s) for patent or inventor's certificate or any PCT application(s) designating at least one country other than the United States filed by me on the same subject matter and having a filing date before that of the application(s) from which priority is claimed:

Country	Number	Filing Date	Priority Claimed	
			Yes	No
FRANCE	96 01309	February 02, 1996	X	

I hereby claim benefit under Section 120 of Title 35 of the United States Code of any United States application(s) or PCT application(s) designating the United States identified below and, insofar as the subject matter of each of the claims of this application is not disclosed in said prior application(s) in the manner provided by the first paragraph of Section 112 of Title 35 of the United States Code, I acknowledge my duty to disclose material information of which I am aware as defined in Section 1.56 of Title 37 of the Code of Federal Regulations which occurred between the filing date of the prior application(s) and the national or PCT filing date of this application:

Application Serial No.Filing DateStatus

I hereby appoint Mary P. Bauman, Reg. No. 31,926; Michael D. Alexander, Reg. No. 36,080; and Paul E. Dupont, Reg. No. 27,438, or any of them my attorneys or agents with full power of substitution and revocation to prosecute this application and to transact all business in the Patent and Trademark Office connected therewith.

SEND CORRESPONDENCE TO:

DIRECT TELEPHONE CALLS TO:

Patent Department  
Sanofi Pharmaceuticals, Inc.  
9 Great Valley Parkway  
P.O. Box 3026  
Malvern, PA 19355

MICHAEL D. ALEXANDERTelephone No. (610) 889-8802

I hereby declare that all statements made herein and in the above-identified specification of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.

Full name of first joint inventor

Daniel CAPUT

Inventor's signature

Daniel CAPUT

Date

16-7-98

Residence

La Bousquière, 31290 Avignonet Lauragais, France

Post Office Address

La Bousquière, 31290 Avignonet Lauragais, France

Citizenship

French

200 Full name of second joint inventor Pascual FERRARA  
Inventor's signature Pascual Ferrara Date 16-7-98  
Residence Libouille Saint-Assiscle, 31290 Avignonet Lauragais, France FIX  
Post Office Address Libouille Saint-Assiscle, 31290 Avignonet Lauragais, France  
Citizenship France

300 Full name of third joint inventor Ahmed Mourad KAGHAD  
Inventor's signature A. Mourad KAGHAD Date 17-7-98  
Residence 5, rue de la Poste, 31450 Montgiscard, France FIX  
Post Office Address 5, rue de la Poste, 31450 Montgiscard, France  
Citizenship French